Article

Current Biology

Harm to Others Acts as a Negative Reinforcer in Rats

Highlights

- Independently of sex and familiarity, rats avoid actions harming a conspecific
- Prior experience with footshocks increases harm aversion
- Rats show large individual variability in harm aversion
- Anterior cingulate cortex deactivation abolishes harm aversion

Authors

Julen Hernandez-Lallement, Augustine Triumph Attah, Efe Soyman, Cindy M. Pinhal, Valeria Gazzola, Christian Keysers

Correspondence

julien.her@gmail.com (J.H.-L.), c.keysers@nin.knaw.nl (C.K.)

In Brief

Hernandez-Lallement et al. show that male and female rats avoid actions that harm familiar and unfamiliar conspecifics. Although prior experience with pain increases this effect, harm aversion is abolished by high cost to help. Deactivation of anterior cingulate cortex area 24, associated with empathy for pain in humans, abolishes this harm aversion.

Hernandez-Lallement et al., 2020, Current Biology *30*, 949–961
March 23, 2020
https://doi.org/10.1016/j.cub.2020.01.017

CellPress

Current Biology
Article

Harm to Others Acts as a Negative Reinforcer in Rats

Julen Hernandez-Lallement,[1,*] Augustine Triumph Attah,[1] Efe Soyman,[1] Cindy M. Pinhal,[1] Valeria Gazzola,[1,2,3] and Christian Keysers[1,2,3,4,*]
[1]Social Brain Lab, Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences, Meibergdreef 47, 1105 Amsterdam, the Netherlands
[2]Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 166, 1018 Amsterdam, the Netherlands
[3]Senior author
[4]Lead Contact
*Correspondence: julien.her@gmail.com (J.H.-L.), c.keysers@nin.knaw.nl (C.K.)
https://doi.org/10.1016/j.cub.2020.01.017

SUMMARY

Empathy, the ability to share another individual's emotional state and/or experience, has been suggested to be a source of prosocial motivation by attributing negative value to actions that harm others. The neural underpinnings and evolution of such harm aversion remain poorly understood. Here, we characterize an animal model of harm aversion in which a rat can choose between two levers providing equal amounts of food but one additionally delivering a footshock to a neighboring rat. We find that independently of sex and familiarity, rats reduce their usage of the preferred lever when it causes harm to a conspecific, displaying an individually varying degree of harm aversion. Prior experience with pain increases this effect. In additional experiments, we show that rats reduce the usage of the harm-inducing lever when it delivers twice, but not thrice, the number of pellets than the no-harm lever, setting boundaries on the magnitude of harm aversion. Finally, we show that pharmacological deactivation of the anterior cingulate cortex, a region we have shown to be essential for emotional contagion, reduces harm aversion while leaving behavioral flexibility unaffected. This model of harm aversion might help shed light onto the neural basis of psychiatric disorders characterized by reduced harm aversion, including psychopathy and conduct disorders with reduced empathy, and provides an assay for the development of pharmacological treatments of such disorders.

INTRODUCTION

Learning to avoid actions that harm others is an important aspect of human development [1], and callousness to others' harm is a hallmark of antisocial psychiatric disorders, including psychopathy and conduct disorder with reduced empathy [2]. What could motivate humans and other animals to refrain from harming others? An influential theory posits that vicarious emotions (i.e., emotions felt by a witness, in the stead of the witnessed individual), including emotional contagion and empathy, trigger harm aversion [3]. Put simply, harming other people is unpleasant, because we vicariously share the pain we inflict. Accordingly, it has been argued that psychiatric disorders characterized by antisocial behavior [2, 4] might stem from malfunctioning or biased vicarious emotions [5, 6].

An increasing number of studies show that rodents display affective reactions to the distress of conspecifics [7–16]. These reactions are observed as increased freezing and modulation of pain sensitivity of the witness while attending to the other conspecific in pain [7–11, 13] or when the witness is re-exposed to cues associated with the other's pain [17, 18]. Recent studies in rats identified emotional mirror neurons in the anterior cingulate cortex (ACC; area 24 in particular) [19, 20], which respond to the observer experiencing pain and to witnessing a conspecific's distress. Reducing activity in the ACC reduces emotional contagion [7, 20]. However, in these paradigms, the observing rat is not the cause of the witnessed pain, and whether vicarious activity in area 24 is associated with harm aversion thus remains unclear.

Inspired by classic studies, here, we refine a paradigm to study instrumental harm aversion in rats. A rat called the ''actor'' can press one of two levers for sucrose pellets. After a baseline phase revealing the rat's preference for one of the levers, we pair this preferred lever with a shock to a second rat (''victim''), located in an adjacent compartment (Figure 1). We then measure how much actors switch away from the preferred lever as a behavioral index of harm aversion.

We show that (1) male and female Sprague-Dawley rats switch significantly away from the shock-delivering lever, (2) this effect is stronger in shock-pre-exposed actors, and (3) deactivating the ACC reduces this effect. By altering the timing of shock delivery, we show that contingency between lever pressing and shock delivery is essential. By varying the reward value of the levers, we show rats switch from an easier to a harder lever and from one that provides two pellets to one that provides one pellet to prevent harm to another. However, rats were unwilling to switch from a lever that provides three pellets to one that provides one pellet. We additionally report and explore substantial individual differences in switching across rats.

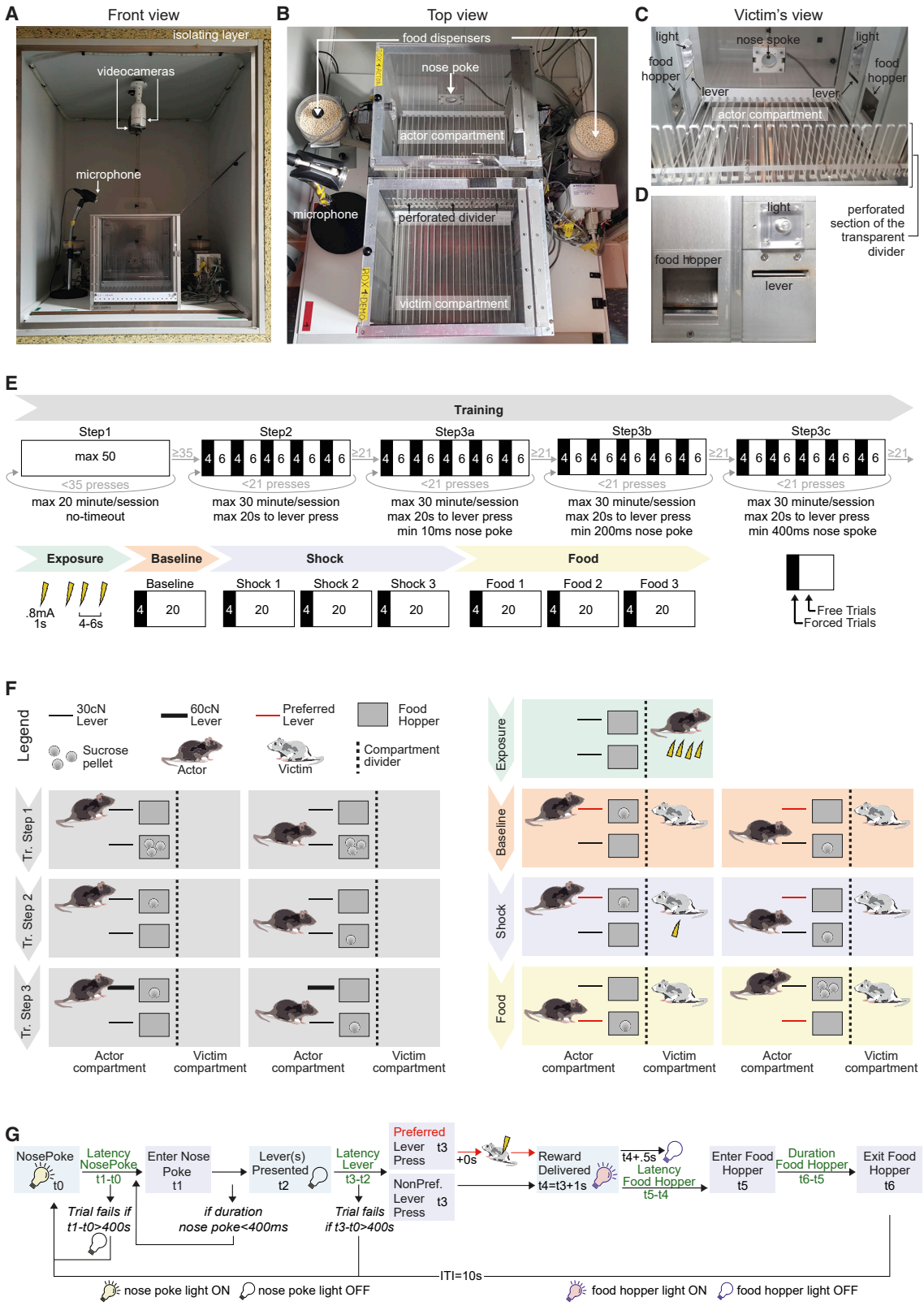


Figure 1. Experimental Procedures

(A) Photo of the cabinet in which the experimental setup was isolated.

(B) Top view of the two-compartment setup.

(legend continued on next page)

Table 1. Core Experimental Conditions

	ContingentHarm		NoHarm		RandomHarm	
	n = 24 (12♂, 12♀)		Nn= 14, all ♂		n = 8, all ♂	
	Shock to Actor	Shock to Victim	Shock to Actor	Shock to Victim	Shock to Actor	Shock to Victim
Exposure	yes	/	yes	/	yes	/
Baseline	no	no	no	no	no	no
Shock	no	yes	no	no	no	yes
Food	no	no	no	no	no	no

For each experimental condition (columns), the table specifies who got electrical shocks during the exposure, baseline, and shock sessions (yes) and who did not (no). A "/" indicates that the victim was not present during the exposure session. Sample size (n) reflects the number of actors included in the behavioral analyses. ♂, male; ♀, female.

RESULTS

Rats Switch Away from a Lever that Triggers Footshocks to a Conspecific

We first compare the behavior of rats in three main conditions: ContingentHarm; NoHarm; and RandomHarm (Table 1). In all three conditions, an actor was trained to press one of two levers for one sucrose pellet in the actor compartment (Figures 1A–1F). One lever required ~60 cN (~60 g) of force to be pressed, although the other required ~30 cN (~30 g), with the harder-to-press side randomized across animals. After initial training alone, all actors were exposed to 4 footshocks (exposure; Figures 1E and 1F) in the adjacent victim's compartment to maximize emotional contagion [9, 19, 21]. Actors were then placed back into their actor compartment and performed 24 trials of lever pressing with their cage mate in the victim compartment (baseline).

These 24 trials started with 4 forced trials (2 for each lever; pseudo-randomized) to force actors to sample both options, followed by 20 free choice trials to measure baseline lever preference (Figure 1E). In the ContingentHarm condition, on the 3 days following baseline (Shock1, Shock2, and Shock3 sessions; Figure 1E), the actor performed 24 trials of the same task each day (4 forced + 20 free choice), similar to baseline trials, except that pressing the lever preferred during baseline triggered a footshock (0.8 mA; 1 s) to the victim in the adjacent compartment. In this condition, we had two groups: male (ContingentHarm ♂) and female pairs (ContingentHarm ♀). We compared this condition against a NoHarm control condition, in which pressing either lever never delivered a shock to the victim to control for spontaneous changes in preference. Finally, we created a RandomHarm control condition, in which the victim is exposed to the same shocks that triggered strong switching in the ContingentHarm condition but were administered independently of the choices of the actor. For this RandomHarm condition, we identified the 8 actors from the ContingentHarm condition (from all 24 animals) that showed the strongest switching away from the shock lever. For each, we recorded the sequence of shock and no-shock trials to the victim. In the RandomHarm condition, each victim then received the sequence of shocks from one of the switchers from the ContingentHarm condition, independently of what lever was pressed by the actor. Crucially, to break the action-outcome contingency, shocks were delayed randomly by 3–8 s after actors exited the food receptacle, i.e., before the start of the following trial.

At the group level, we compared preference changes from baseline to shock sessions across conditions. A 4-group$_{(\text{ContingentHarm}♂,\ \text{ContingentHarm}♀,\ \text{NoHarm}♂,\ \text{RandomHarm}♂)}$ × 4-session$_{(\text{baseline},\ \text{Shock1},\ \text{Shock2},\ \text{Shock3})}$ repeated-measures ANOVA revealed a significant effect of session ($F_{(3,123)} = 7.34$; $p < 0.001$; $\eta^2 = 0.15$; $BF_{incl} = 433$) and session × group interaction ($F_{(9,123)} = 1.93$; $p = 0.05$; $\eta^2 = 0.12$; $BF_{incl} = 3$). We first concentrate on male actors, for which we have three groups (ContingentHarm♂, NoHarm♂, and RandomHarms♂), which showed similar preferences at baseline (i.e., comparable preference for the future shock lever; ANOVA; $F_{(2,31)} = 1.29$; $p = 0.289$; $BF_{incl} = 0.46$). From baseline to all shock sessions, ContingentHarm male actors showed the expected decrease in shock lever pressing, with their preference for the shock lever lower than the NoHarm and RandomHarm control groups in all shock sessions (even if regressing out differences in baseline preference; Figure 2A). Actors in the male ContingentHarm group thus shifted significantly away from a lever that causes shocks to a conspecific, and this was not simply due to the distress of the victim (which was matched, i.e., no significant differences in the amount of freezing and ultrasonic vocalizations (USVs), across ContingentHarm and RandomHarm groups; Figure S1) but to the

(C) View of the actor's compartment through the perforated transparent divider as seen from the victim's compartment.
(D) Close up of the left wall with one of the two levers and food hoppers available to the actor.
(E) Experimental timeline from the training to the end of the experiment. The numbers within each block indicate the number of lever presses (in training step 1) or trials (in training steps 2 to 3, baseline, shock, and food sessions) maximally allowed (training step 1) or required (training steps 2 and 3, baseline, shock, and food sessions). Training steps 1–3 were repeated until 70% of the maximum number of possible lever presses per session in the free trials was reached (i.e., 35 in Step1 and 21 in Step2–3, gray numbers). A time out for the lever press was introduced in step 2. An additional time out based on the duration of the nose poke was introduced in step 3. This duration went from a minimum of 10 ms to a minimum of 400 ms in three separate sessions.
(F) Design of training, exposure, baseline, shock, and food sessions. During step 1 of the training, the animal is free to press any lever and each lever press delivers three pellets. In step 2, the animal is exposed to 5 blocks, each starting with four forced trials (only one lever at the time is presented, which has to be pressed within 20 s) and finishing with 6 free trials (both levers presented and one has to be pressed within 20 s). Only one pellet is given at each lever press. In step 3, a difference in strength necessary to operate one of the levers is introduced and nose spoke is required in order to initiate a trial. Again, animals are exposed to 50 trials each session, with the same number of forced and free trials as in step 2. During exposure, the actor receives 4 shocks alone in the victim compartment. During baseline, pressing one lever (left column) or the other (right column) leads to one pellet for the actor. The lever preferred during baseline then additionally triggers a shock to the victim during the 3 days of the shock phase. During food sessions, shocks are no longer delivered, but the non-preferred lever from shock session 3 now leads to 3 pellets. Note that the food session is only present in some conditions (Tables 1 and 2).
(G) Trial structure for the shock session. Trial structure for baseline and food is identical to shock except that shocks are never delivered to the victim, and three pellets are delivered to the actor when the non-preferred lever is pressed in the food session. Light blue boxes denote events and lavender boxes the rat's responses. Latency nosepoke and latency lever are used as criteria to determine valid trials, although latency and duration food hopper are dependent variables reported in Figures 3 and S2.

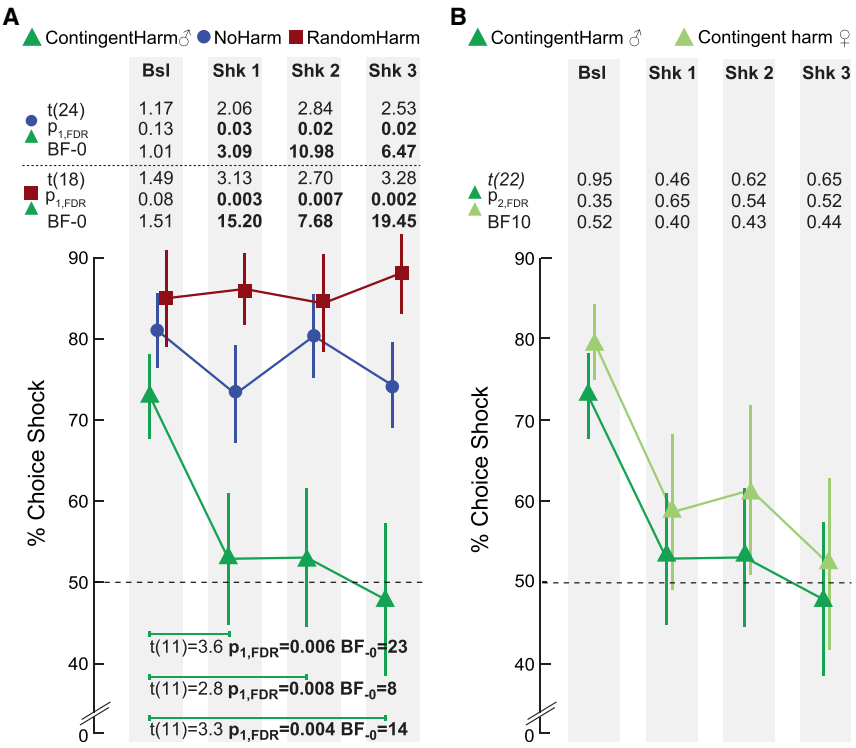


Figure 2. Harm Aversion in Rats

(A) Percent choice for shock lever across sessions (baseline, Shock1, Shock2, and Shock3) and conditions (ContingentHarm, NoHarm, and RandomHarm). The numbers above the graph specify the t-statistic (t), the one-tailed false discovery rate (FDR) corrected p (p1,FDR) and the one-tailed Bayes factor (BF-0) for the comparisons between the two conditions indicated leftmost, separately for each session. In the lowest part of the graph, statistics for pairwise comparison between each shock session and the baseline are shown. Values in bold are significant. Bsl, baseline; Shk1–3, shock sessions 1–3. Gray rectangles on background help visually discriminate the sessions. Data are presented as mean ± SEM.

(B) Same as in (A) but comparing female and male ContigentHarm groups and reporting two-tailed t test (p2,FDR) and two-tailed Bayes factor (BF10). See also Figures S1 and S4.

contingency between the actions of the actor and the reactions of the victim.

Male and females did not differ in their change in preference across sessions (Figure 2B; session × gender: $F_{(3,63)} = 0.21$; $p = 0.89$; $\eta^2 = 0.01$; $BF_{incl} = 0.20$), with both showing a significant main effect of session when analyzed individually (female: $F_{(3,30)} = 5.9$, $p < 0.003$, $BF_{incl} = 14.5$; Male: $F_{(3,33)} = 7.8$, $p < 0.001$, $BF_{incl} = 64$). For all subsequent analyses looking at the change of preference across sessions, we thus pool males and females into one single ContingentHarm condition (n = 24 actors). The Bayes factor for including a main effect of gender, however, was anecdotal ($BF_{incl} = 0.44$).

To prevent harm to their victim, actors could stop pressing any lever instead of switching to the no-shock lever. Across our three groups, all animals performed all their baseline trials. In the ContingentHarm shock sessions, six animals failed to press any lever within the 400 s allowed (missing 1, 1, 2, 10, 20, and 32 out of the 60 free choice trials over the 3 shock sessions, respectively). In contrast, all animals in the NoHarm condition performed all their 60 free choices over all sessions, and only one in the RandomHarm condition missed one trial. This illustrates witnessing contingent shocks to another rat can motivate agents to stop pressing levers altogether. However, given that, over all ContingentHarm animals, 95% of free trials were performed, we concentrate on the shift away from the shock lever as our dependent measure.

Rats Show Substantial Individual Differences in Switching

To quantify switching at the individual level, we computed a switching index (SI),

$$SI = \frac{S_{baseline} - S_{shock}}{S_{baseline} + S_{shock}},$$

where $S_{baseline}$ is the proportion of shock lever presses during baseline and S_{shock} the average proportion of shock lever presses over all shock sessions (see STAR Methods). Positive SIs reflect switching away from the shock lever and SI = 1 maximum possible switch given an individual's baseline preference (Figure 3D). In the ContingentHarm condition, some animals showed substantial preference changes in shock sessions, and others remained indifferent. A permutation test revealed n = 9 actors (i.e., 38%) in the ContingentHarm condition (n = 4 males and n = 5 females) showed a significant switch (at $p < 0.05$; green solid circles in Figure 3A; hereafter referred to as "switchers"). A binomial test showed that 9 out of 24 switchers are not explained by chance (binomial; n = 24; alpha = 0.05; $p = 10^{-6}$). These switchers found across males and females showed a decrease between 25% and 80% from baseline. Switching rates were within chance level in the NoHarm (n = 1 significant switcher; blue colored in circle; binomial; p = 0.36) and absent in the RandomHarm condition (n = 0 significant switchers). A χ^2 square test revealed the ContingentHarm condition had more switchers than the NoHarm condition ($\chi^2 = 4.20$; $p = 0.04$) and the RandomHarm condition ($\chi^2 = 4.17$; $p = 0.04$). Figure 3B shows the lever choices session per session for each ContingentHarm actor and the distribution of changes across sessions. For switchers, most changes in lever choice occurred in the first shock session, with little change occurring in the subsequent sessions.

To explore what may determine these differences in switching in the ContingentHarm condition, we extracted a number of variables from the behavior of the actors and victims and examined which could predict the SI (Table 2; Figure S2). To limit multiple comparisons, we focused on a limited number of variables that are meant to assess distress, attention, the ability to press the levers, and behavioral flexibility. Behavioral flexibility was assessed in two ways: (1) how much the preference for the lever that will later be paired with shocks changed from step 3c of the training session to baseline (Figure 1E) and (2) how much

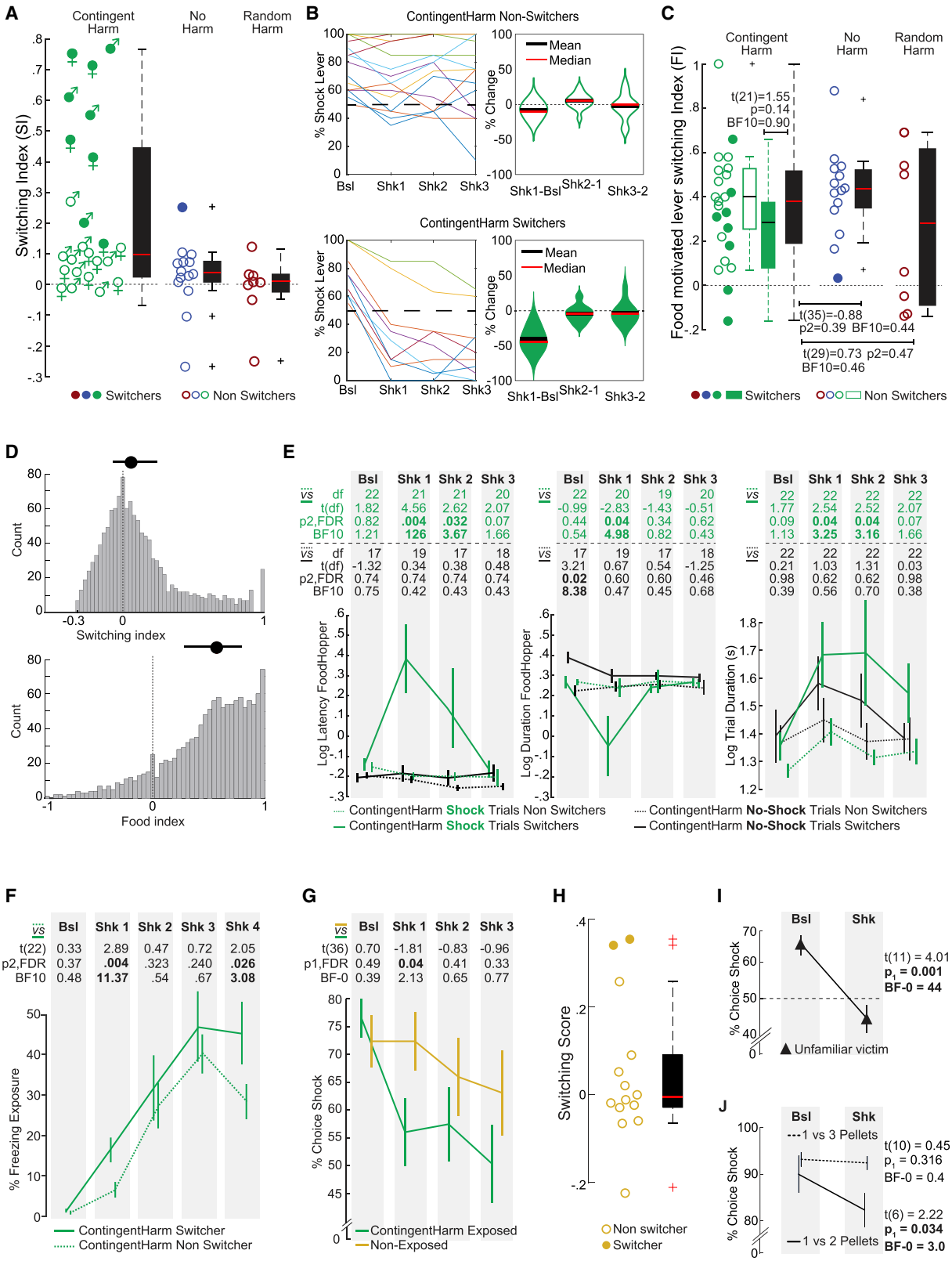


Figure 3. Individual Variability

(A) Switching index (SI) for the different conditions. The dots represent the SI values of each rat separately for the three main conditions. Filled dots indicate the rats that switched to the non-preferred lever. Boxplot of each distribution shows the median (red bar), outliers (crosses), and the 25% and 75% percentile values.

(legend continued on next page)

actors switched lever preferences after the shock session for food reward. The latter was measured after the end of the third shock session by turning off shock delivery, identifying which lever was less preferred, and baiting that lever with 3 pellets, in contrast to one pellet delivered by pressing the preferred lever (food session; Figures 1E and 1F; Table 1). We computed individual food indices (FIs) (Figures 3C and 3D), which quantified the change of lever choice from the last shock session across the three successive food sessions. ContingentHarm animals did not show significant differences in FI from the NoHarm and RandomHarm animals (Figure 3C), and switchers and non-switcher animals showed comparable FI in the ContingentHarm condition (Figure 3C), suggesting that non-switchers switch as much as switchers for rewards, but not for shocks to others.

To relate all these measures to SI (which is not normally distributed over the entire group), we used Kendall's Tau rank order correlation as the measure of association. Table 2 and Figure S2 show these variables ranked by the evidence (Bayes factor) for an association. Focusing on associations with a $BF_{10} > 3$ (dark red lines in Figure S2) shows that animals that switched more spent *less* time in the food hopper and took *longer* to enter the food hopper after trials in which the victim received a shock, leading to longer overall trial duration. Figure 3E illustrates that this effect is visible specifically in the Shock1 session, where switchers, but not non-switchers, delay their entry and accelerate their exit from the food hopper specifically on trials in which they delivered a shock to the other animal. This was confirmed by an ANOVA that revealed a $\text{session}_{(\text{baseline, Shock1, Shock2, Sochk3})} \times \text{trial}_{(\text{shock lever, no-shock lever})} \times \text{type}_{(\text{switchers, non-switchers})}$ interaction (significantly for log latency $F_{(3,36)} = 5.9$, $p = 0.002$, $BF_{incl} = 73$ and a trend for log duration $F_{(3,33)} = 2.37$, $p = 0.08$, $BF_{incl} = 2.7$). A similar effect was not apparent in the NoHarm or RandomHarm conditions (Figure S3). As a result, switcher rats also took longer to perform trials (Figure 3E). This suggests witnessing the victim receive shocks interfered with the food-directed action of switchers, but not non-switchers. We also observed that dyads with more switching had victims that spent *less* time close to the divider. In contrast, variables that might have captured differences in distress signals (freezing, 22-kHz USV emissions, and loudness of pain squeaks) failed to reveal robust associations with switching (Table 2). The same was true for weight and our measures of behavioral flexibility, as measured by changes in lever choice across training and baseline or in response to food rewards.

Prior experience with footshocks increases the sensitivity of rodents to witnessing footshocks in others [9, 16, 19, 21, 22]. Does prior experience also influence switching in our paradigm? During the exposure sessions, animals increased freezing (Figure 3F). We found a hint toward higher average freezing during the shock epochs of the pre-exposure in switchers compared to non-switchers (thick versus dotted lines in Figure 3F; $t_{(22)} = 1.55$; $p_2 = 0.135$; $BF_{10} = 0.88$). A similar trend was observed in actors tested with unfamiliar victims ($t_{(10)} = 1.8$; $p_2 = 0.102$; $BF_{10} = 1.2$). However, these data remain inconclusive. To further test the importance of prior exposure, we tested a new group that followed the same procedure as the ContingentHarm condition, except that the actors received no shocks during the exposure session (NonExposed; Table 3). We observed no significant effect of session on lever preference in NonExposed animals (Figure 3G; one-way ANOVA; $F_{(3,39)} = 1.45$; $p = 0.242$; $BF_{10} = 0.38$), and a repeated measures ANOVA (rmANOVA) using both ContingentHarm and non-exposed conditions revealed a significant session ($F_{(3,55)} = 8.71$; $p < 0.001$; $\eta^2 = 0.20$; $BF_{incl} = 10{,}819$) and session × condition interaction ($F_{(3,105)} = 3.51$; $p = 0.018$; $\eta^2 = 0.09$; $BF_{incl} = 5.9$). Although baseline preference levels were comparable across both conditions, ContingentHarm animals showed significantly lower preference for the shock lever during the first shock session compared to the NonExposed condition (Figure 3G; pairwise comparisons). This difference becomes nonsignificant in the subsequent shock sessions. Switching was within chance levels in the NonExposed condition ($n = 2$ significant switcher; Figure 3H, filled yellow dots; binomial; $p = 0.12$). Together, these analyses show that prior fear experience primes rats to a higher sensitivity to other's pain.

In summary, we thus identified two main factors associated with individual differences in switching. First, prior experience with footshocks increases switching. Second, animals that switch more show a stronger reaction to the shocks of the victim: they delay entering the food hopper, spend less time in the food hopper, and take longer to perform trials. In contrast,

(B) Individual lever preference (left) and change in lever preference (right) as a function of session, for switcher (top) and non-switcher (bottom) in the ContingentHarm group.
(C) Food index for the different conditions. Boxplot of each distribution shows the median (red bar) and outliers (crosses). For the ContingentHarm condition, the boxplots have been computed separately for the switchers (green filling), the non-switchers (green contour), and the whole group (black filling). For the NoHarm and RandomHarm, only the group results are presented (black filling) because there are insufficient switchers. See also Figure S2 for related results and data.csv for the choice data.
(D) Histograms of SI and FI values obtained when computing the indices based on randomly drawn Bsl, Shk1…Shk3 values from uniform distributions. Black dot and line: distribution's median and the 25% and 75% percentile values.
(E) Log-transformed latency to enter the food hopper, duration in the food hopper, as well as average trial duration for switchers and non-switchers and shock and no-shock trials in the ContingentHarm condition. See Figure S3 for similar data for the NoShock and RandomShock conditions.
(F) Percentage freezing during the shock exposure sessions for the animals that will become switchers (thick line) and non-switchers (dotted line). The percent freezing quantifies the percentage freezing in the following inter-shock interval.
(G) Percent shock lever presses in ContingentHarm for the exposed (green) and non-exposed group (yellow) across sessions.
(H) Switching score distribution for non-exposed animals.
(I) Average percent shock lever presses during baseline and shock sessions (i.e., average over 2 baseline sessions and 2 shock sessions) for the actors paired with unfamiliar victims.
(J) Percent shock lever presses for baseline and average of 3 shock sessions for 1 versus 2 pellets and 1 versus 3 pellets.
Data in (E–G), (I), and (J) are mean ± SEM. Numbers in bold are significant; BF, Bayes factor; p1 and p2 indicate one-tailed and two-tailed testing, respectively. FDR, p values corrected using false discovery rate for 4 sessions; df, degree of freedom, which is lower in sessions in which some animals never chose the no-shock or the shock lever; See data.csv at https://osf.io/65j3g/ for the data that went into (C). See also Figures S2 and S3 and Tables S1, S2, and S3.

Table 2. Behavioral Correlates of Switching

	tau	Lower 95% CI	Upper 95% CI	BF_{10}	p2
Difference in log time spent in the food hopper (Shock–NoShock in session Shock1)	−0.55	−0.75	−0.18	46.46[a]	0.001[a]
Difference in log latency to enter food hopper (Shock–NoShock in Shock1)	0.46	0.12	0.66	14.32[a]	0.004[a]
Increase victim time spent close to divider (Shock–NoShock in Shock1)	−0.40	−0.61	−0.10	9.28[a]	0.006[a]
Increase in average trial duration (Shock1–baseline)	0.39	0.08	0.60	7.57[a]	0.008[a]
Change in lever preference for food (food index)	−0.30	−0.53	0.00	1.86	0.045[a]
Spontaneous changes in shock lever preference (last training–baseline)	−0.26	−0.49	0.03	1.11	0.082
Difference in weight (actor–victim)	−0.23	−0.46	0.06	0.85	0.118
Increase in actor freezing (Shock1–baseline)	0.16	−0.12	0.40	0.46	0.295
Increase in victim freezing (Shock1–baseline)	0.14	−0.14	0.39	0.41	0.333
Weight actor	−0.11	−0.36	0.16	0.34	0.456
Weight victim	0.05	−0.22	0.30	0.28[a]	0.747
Increase actor time spent close to divider (shock–NoShock in Shock1)	0.05	−0.22	0.30	0.28[a]	0.747
Squeak loudness (power shock–NoShock in Shock1)	−0.05	−0.30	0.22	0.28[a]	0.747
Increase in 22 kHz USV (Shock1–baseline)	0.00	−0.26	0.26	0.26[a]	0.980

Variables ranked based on decreasing evidence of correlation (BF10) using the rank order correlation Kendall Tau. The horizontal lines separate variables based on whether there is (1) evidence for the presence of an association (top cases, $BF_{10} > 3$), (2) inconclusive evidence (middle, $0.33 < BF10 < 3$), or (3) evidence for the absence of an association ($BF10 < 0.33$, bottom). See also Figure S2. BF10, Bayes factor in favor of the presence of a correlation; CI, confidence interval; P2, two tailed frequentist probability for Tau = 0; Tau, Kendall's tau.

[a]Significant either based on BF (using 3 and 1/3 as critical values) or based on $p < 0.05$ threshold

quantifications of the behavior of the victim appear not to predict switching. It thus appears, within our paradigm, as though the main determinant of individual differences stems from the actor, not the victim: among the variables we quantified, it is how the actor reacts to the victim and prior shock experience with shock, not how the victim reacts to the shocks, that are most associated with switching.

Familiarity with the Victim Is Not Necessary for Switching

During the piloting phase of the paradigm, we tested actors that were unfamiliar with their victims taken from unrelated cages (unfamiliar victims; Table 3). Actors showed a significant decrease from baseline preference levels also for shocks to these unfamiliar victims (paired one-tail t test; Figure 3I), and 2 out of the 12 actors were detected as significant switchers. This effect was comparable to the one observed in the ContingentHarm animals of our main experiment ($session_{(baseline,\ average\ shock)} \times condition_{(ContingentHarm,\ unfamiliar\ victims)}$; $F_{(1,34)} = 0.20$; $p = 0.66$; $\eta^2 = 0.006$; $BF_{incl} = 0.80$). Accordingly, familiarity is not necessary for reducing lever preference, in line with data showing that rats freeze even when an unfamiliar conspecific gets a shock [16] and free an unfamiliar trapped rat [23]. However, we cannot exclude that familiarity may have a subtle effect on the magnitude of switching, as shown in mice [24–28]. It is important to note that, during the piloting phase of the experiment, actors in the unfamiliar victims condition were also exposed to shocks prior to the experiment but in another context rather than in the victim compartment. Hence, this condition shows that the switch in preference observed in the ContingentHarm condition is not solely due to contextual fear formed during the exposure session in the victim's compartment.

Switching Is Modulated by Cost

To test whether actors would give up food to avoid another's distress, we tested new groups of rats (1vs2Pellets and 1vs3Pellets; Table 3), where levers required the same effort (~30 cN) but differed in rewards from the beginning of the baseline session. In the 1vs2 condition, the shock lever provided n = 2 pellets, although the no-shock lever provided n = 1 pellet. Actors decreased their preference for the 1vs2 option upon association with a shock (paired one-tailed t test; Figure 3J, solid line), and 3 out of 7 actors in this group were detected as significant switchers. To explore whether this switching differed from the ContingentHarm, we computed a repeated-measures ANOVA using ContingentHarm and 1vs2Pellets conditions with 4 sessions each. We found a highly significant main effect of session ($F_{(3,84)} = 7.03$; $p < 0.001$; $\eta^2 = 0.20$; $BF_{incl} = 175{,}403$) but only a trend for an interaction ($F_{(3,84)} = 1.82$; $p = 0.15$; $BF_{incl} = 2.96$). Hence, rats are willing to forgo one sucrose pellet to avoid the victim's distress, but the effect tends to be slightly reduced compared to a difference in effort. The 1vs3 pellet condition, where the levers led to n = 1 versus n = 3 pellets (Figure 3J, dotted line) did not show a significant decrease of preference from baseline (paired one-tailed t test), none of the actors were significant switchers, and a rmANOVA ($2\ groups_{(ContigentHarm,\ 1vs2Pellets)} \times 4\ sessions_{(baseline,\ Shock1,\ Shock2,\ Shock3)}$) shows the effect was significantly smaller than in the ContingentHarm

Table 3. Paradigm Contingencies for Additional Conditions

	Unfamiliar Victims n = 12♂		1vs2Pellets n = 7♂		1vs3Pellets n = 11♂		NonExposed n = 14♂		Muscimol n = 11♂		Saline n = 12♂	
	Shock Actor	Shock Victim	Shock Actor	Shock Victim	Shock Actor	Shock Victim	Shock Actor	Shock Victim	Shock Actor	Shock Victim	Shock Actor	Shock Victim
Exposure	yes	/	yes	/	yes	/	no	/	yes	/	yes	/
Baseline	no	no	no	no	no	no	no	no	no	no	no	no
Shock	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
Food	/	/	/	/	/	/	/	/	no (6)	no (6)	no (7)	no (7)

In all but the food row, "/" indicates that the animal was not present, "yes" that the animal was present and received shocks, and "no" that the animal was present but received no shocks. For the food row, / indicates groups in which the food condition was not run. No (x) indicates that the condition was run in x of the N animals but that no shock was delivered. Sample size (N) reflects the number of actors included in the behavioral analyses.

condition (interaction; $F_{(3,96)} = 5.46$; $p = 0.002$; $BF_{incl} = 148$), suggesting that harm aversion may not be strong enough to counteract high costs.

Prolonged Training Reduces Switching

In our experiments, rats were not required to develop strong and stable preferences for a lever before associating the preferred lever with a shock. However, some animals showed a consistent preference for the same lever over the last training session and the baseline session and showed significant switching (see Figure S4A). To investigate the impact of more pronounced preferences, we used data from two additional groups in which animals (n = 21) were trained to reach stable preference levels >80%, which took on average 480 additional trials compared to ContingentHarm, NoHarm, and RandomHarm conditions. In these "over-trained" animals, we did not find significant switching at the group level (see Figure S4B). These results therefore suggest that the more habitual a behavior, the less it is sensitive to modification by social consequences.

The ACC Is Necessary for Harm Aversion in Rats

Several studies performed in humans [5, 29–31] and rats [7, 19–21] suggest the ACC (including area 24a and b) [32] is recruited during the observation of distress and maps other's pain onto one's own pain circuitry. To test whether the ACC is necessary for the switching in our paradigm, we infused muscimol bilaterally in the ACC in a group of rats (muscimol; Table 3) prior to baseline and shock sessions and compared the choice allocation to a saline-infused group (saline; Figure 4A; Table 2). Infusions were centered at +1.8 mm from the bregma and had an anterior-posterior spread of [+1.95 mm; +1.45 mm] from bregma (muscimol group: n = 9 out of 11; M = 1.76 mm; SD = 0.26; saline group: n = 9 out of 12, M = 1.84; SD = 0.26; Figure 4B), confirming that area 24a (approximately 0.6 mm dorsal to corpus callosum) and area 24b were targeted. The infusion spared midcingulate areas 24', located closer to the bregma [32, 33], as well as deeper area 33 and the cingulum, located postero-ventral to most infusions [34].

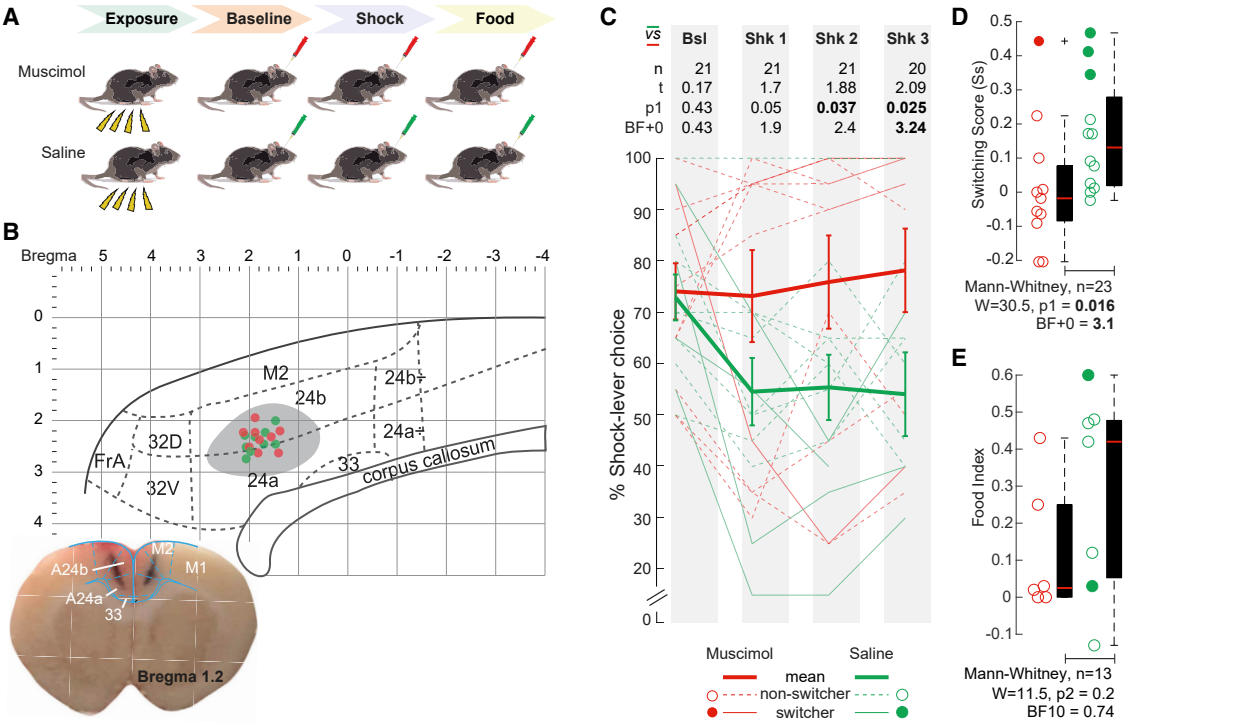
A $2\text{-condition}_{(muscimol,\ saline)} \times 4\text{-session}_{(baseline,\ Shock1,\ Shock2,\ Shock3)}$ rmANOVA revealed a trend for an effect of session ($F_{(3,60)} = 2.32$; $p = 0.08$; $\eta^2 = 0.10$; $BF_{incl} = 0.88$) and a significant session × condition interaction ($F_{(3,60)} = 3.33$; $p = 0.025$; $\eta^2 = 0.14$; $BF_{incl} = 2$). Although baseline preferences for the shock lever were comparable between muscimol and saline, preferences for the shock lever were significantly higher for the muscimol- than saline-infused animals in shock sessions (Figure 4C), and switching scores were significantly reduced in the muscimol compared to the saline group (Mann-Whitney U = 30.5, $p_1 = 0.016$; Mann-Whitney $BF_{+0} = 3.1$; Figure 4D).

Muscimol injection did not appear to alter goal-directed behavior more generally in terms of latency to enter and duration in the food hopper, percentage of time spent close to the divider, and the amount of freezing (Figure 4F). We also started including food sessions in later animals (saline n = 7 and muscimol n = 6; Table 3), continuing the injection of the respective drug during those sessions. In that small subsample, FIs were not significantly different between muscimol and saline conditions (Figure 4E), but the Bayesian analysis suggests a larger sample is necessary to exclude small effects on flexibility. If we repeat the ANOVA on the subsample of n = 7+6 animals for which we performed food session, we still find that switching is reduced after ACC deactivation ($2\ condition_{(muscimol,\ saline)} \times 4\ session_{(baseline,\ Shock1,\ Shock2,\ Shock3)}$ interaction; $F_{(3,33)} = 5.6$; $p = 0.003$; $BF_{incl} = 12.7$).

DISCUSSION

By characterizing the willingness of rats to reduce the use of a lever that triggers pain to another conspecific, we provide evidence that rats display a contingency-dependent harm aversion, which we show to be influenced by ACC deactivation. Despite substantial individual variability, harm aversion was replicated at the group level in 4 separate groups of animals that were tested with levers differing in effort (male ContingentHarm, female ContingentHarm, unfamiliar victims, and saline). Significant switching away from a lever harming another rat was also replicated in a condition in which switching involved using a lever delivering one instead of a lever delivering two pellets (2 versus 1 pellet), although this effect was statistically weaker. The willingness to switch was no longer significant when the difference in value across levers was too high (3 versus 1 pellet) or when animals were overtrained (Figure S4).

In accord with the original study of Greene [37], we found substantial individual differences across our rats, with only a subset showing strong switching. Actors that switched more were found to be characterized by the fact that they delayed and shortened their reward consumption following shocks to the victim and oriented more toward the victim in shock trials. Also, an actor's prior experience with shocks increases



(legend on next page)

switching, in line with prior studies showing that prior experience increases sensitivity to others' pain as measured by vicarious freezing [9, 16, 19, 21, 22]. The behavior of the victim, in terms of USVs, pain squeaks, and freezing, in contrast, was not robustly associated with switching, but victims that decreased their time spent close to the divider (due to shock-induced behavioral activity) were paired with actors with higher switching scores. This suggests the possibility of some association between the behavior of the victim and the actors' harm aversion levels. Such individual variability could be valuable to shed light on the origin of individual variance in human harm aversion and merits further attention to identify the signals from the victim that are necessary and sufficient for switching. This is particularly true given that rodent models of the kind of disrespect for other people's well-being encountered in human antisocial behavior are so far lacking [38]. Although, in the majority of actors, harm aversion manifested as a willingness to switch to a less preferred lever, a small number of animals in the ContingentHarm group stopped pressing levers altogether, thereby also preventing shocks to the victim at a cost of up to 60 sucrose pellets.

Additionally, we find sex does not modulate harm aversion. This is in agreement with two studies showing a lack of sex difference in emotional contagion [39] and discrimination [15] but apparent contrast to a small number of studies that reported sex effects on vicarious responses in mice [25, 26, 40] and rats [41, 42], pointing toward a growing awareness that the specific output behavior measured can dramatically alter sex differences [43]. Second, we found no effect of familiarity on harm aversion. This finding is in line with a number of studies finding that familiarity does not influence vicarious freezing and fear transmission in rats [16, 44] or emotion discrimination and pain transmission in mice [14, 28]. Third, we find that the contingency between the actor's actions and the victim's distress is essential to trigger switching: the same number of shocks to the victim without contingency (RandomHarm) triggered similar levels of distress in the victim but fails to trigger switching in actors. Fourth, we find that most of the changes in lever preference and food consumption timing occurred in the first shock session. Reinforcement learning theories suggest the brain learns from prediction errors to choose those options with the highest expected value [45]. In our paradigm, shock session 1 is where a surprising event—the pain of a conspecific—is introduced. That those rats that later changed their choices most were those that showed the strongest changes in latency and duration of the next action is compatible with a vicarious reinforcement learning: they are perhaps those for whom the reaction of the victim triggered the most salient prediction error and hence expected value reduction and decision changes. By the end of the first session, the outcome was then no longer surprising (as suggested by the normalizing latencies) and choices stabilized. That some switchers already reached very low shock lever preferences in the first shock session (Figure 3B, lower panel) additionally creates a floor effect that reduces preference changes in later sessions.

Our experiment has a number of important limitations to consider. First, we found that deactivating the ACC reduces switching. This demonstrates the potential of our paradigm to reveal the involvement of brain regions in harm aversion. Specifically, recent studies have suggested that both in rats and humans, the pain felt by a conspecific is mapped onto our own pain representation through emotional mirror neurons, located within the ACC [19, 20, 46]. Our deactivation data now suggest that this region may be important to prevent harm to others. However, we injected muscimol throughout all sessions, from baseline to shock3, and we thus cannot pinpoint in what phase of the task the ACC is important. Recording cellular activity during the task and using optogenetic deactivation at particular moments in the task will be essential to pinpoint when and how the ACC is important in harm aversion. Recent studies suggest that the medial prefrontal cortex [15] and the amygdala [14] are recruited during the discrimination of the emotions of conspecifics in mice. Whether these structures together with the ACC are involved in harm aversion in rats should be further explored. Second, one would be inclined to interpret our data as suggesting that switchers are willing to exert twice the force to save shocks to others. However, as observed in previous studies [37], approximately 25% of the actors actually preferred the hard to the easy lever in baseline sessions, and for some, switching thus did not involve additional effort but actually a reduction of effort. Rather than showing a willingness to work harder for others, our data thus show that rats are willing to switch to a less-preferred lever. Third, we show that pre-exposure to shocks potentiates switching. It has been shown that freezing while observing another animal receive shocks is potentiated by prior shock experience [9, 19, 21, 47], but other arousing experiences, such as a forced swim test, do not have the same potentiating effect [22, 48]. Moreover, although

Figure 4. The ACC Is Necessary for Switching

(A) Experimental procedures.

(B) Estimated anterior-posterior and dorsal-ventral coordinates of the infusions on a sagittal representation of the medial surface of the rat brain based on [32, 35] for muscimol (red) and saline (green) animals. Each dot is the average of the coordinates of the right and left cannula tip location in histological slices. Gray shading, estimate of likely spread based on a combination of published data [36] and estimates from our own lab based on similar injections of fluorescent muscimol.

(C) Individual (lighter lines) and group (thicker lines) preferences for shock lever across sessions for saline (green) and muscimol animals (red). Thin dotted lines indicate non-switcher animals; thin solid lines indicate switchers.

(D) Switching scores as a function of group (red, muscimol; green, saline). Solid circles represent switchers.

(E) Food index scores as a function of group (red, muscimol; green, saline). Solid circles represent switchers.

(F) Latency to enter and duration in hopper, percent of time spent close to the divider, and percent of freezing after shock and no shock trials for the muscimol (red) and saline (green) groups. Data are mean ± SEM, but Mann-Whitney U test has been reported as values for latency to reward poke and percent of freezing were not normally distributed. Student's t tests, where applicable, confirmed the non-significance of all differences but time spent close to the divider on shock 3 ($t(15) = 2.5$; $p = 0.025$). Significant numbers in bold. p1 and p2 indicate one-tailed and two-tailed testing, respectively.

Also see data.csv at https://osf.io/65j3g/ for data in (C–F).

frequentist statistics suggest an effect of prior shock experience, the Bayes factor remains within a range where caution should be used in the interpretation of the data. Quantifying switching in our paradigm following a forced-swim test instead of footshock exposure would be necessary to explore whether switching also specifically depends on prior experience with a stimulus similar to that of the victim or, less specifically, on any heightened state of arousal or fear that sensitizes actors to any stimulus that could signal danger. Fourth, our unfamiliar group was tested with prior shock exposure in an environment that differed from the test environment although the rats in the main experiments received prior exposure in the test environment itself. That switching was significant at the group level in both cases shows that it is robust against changes in familiarity and context but makes it difficult to isolate the effect of either variable precisely.

Finally, it is important to specify that our data do not show that rats are altruistic in the sense of acting with the intention to benefit someone else. The human literature has introduced a distinction between two motivations to help. Some participants help others because seeing them suffer creates an aversive state called personal distress that participants then try to selfishly reduce by helping [49]. Other participants are more altruistic and help even if they do not have to witness the suffering of the victim, suggesting a more altruistic, truly other-regarding motivation [49]. Our design does not allow us to distinguish these options, but a parsimonious, selfish explanation could suffice to explain our effects: pressing the shock lever triggers reactions in the victim, which, via association with the actor's prior shock experience, can trigger an aversive state and/or fear for the actor's own safety that it tries to avoid by switching to the other lever—or more rarely, by stopping to press levers altogether. In this view, harm aversion may not primarily be an altruistic motive to prevent pain to another rat but a more selfish motive to avoid an unpleasant personal state triggered by the signals emitted by the other rat—a less noble but perhaps equally effective motive. Indeed, rats can be motivated to switch their lever preference also against a panoply of non-social stimuli, including loud noises or bright light [50, 51]. By contrasting social and non-social stimuli, as has been done for vicarious freezing [20], an interesting question for future neuroscience research will be to investigate what brain structures may be specifically involved in modulating behavior based on the pain of others.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Subjects
  - Sample size calculation
- METHOD DETAILS
  - Experimental setup
  - Experimental procedures
  - Surgery and cannulation
  - Humane endpoints
  - Infusion of saline and muscimol
  - Histology
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis of lever presses
  - Switching index
  - Food index
  - Statistical analysis
  - Additional behavioral analysis
  - Audio analysis
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.01.017.

A video abstract is available at https://doi.org/10.1016/j.cub.2020.01.017#mmc3.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 745885 to J.H.-L. and from the Dutch Research Council (NWO) VICI grant (453-15-009) to C.K. and VIDI grant (452-14-015) to V.G. We thank Steven Voges for valuable comments on the manuscript. We thank Susan van der Boogard for her help with the data acquisition, as well as Laura Stolk for her assistance in histological verifications in cannulas implantation. We thank the Mechatronics Department of the Netherlands Institute for Neuroscience for their help in building the experimental setups. We thank animal caretakers of the Netherlands Institute for Neuroscience for valuable support with animal care.

AUTHOR CONTRIBUTIONS

J.H.-L. acquired funding, conceived the study, conducted all experiments, analyzed the data, and wrote the manuscript. A.T.A. acquired data for the ACC deactivation and the effect of prior exposure, analyzed the data for these sections, and provided comments on the manuscript. E.S. analyzed audio data and provided comments on the manuscript. C.M.P. analyzed audio data, performed the histological verification of cannula placement, and provided comments on the manuscript. V.G. acquired funding, conceived the study, analyzed some of the data, and wrote the manuscript. C.K. acquired funding, conceived the study, analyzed some of the data, and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 8, 2019
Revised: September 11, 2019
Accepted: January 7, 2020
Published: March 5, 2020

REFERENCES

1. Wilson, D.S. (2015). Altruism in everyday life. Does Altruism Exist?: Culture, Genes, and the Welfare of Others (Yale University Press), pp. 117–129.
2. Blair, R.J.R. (2013). The neurobiology of psychopathic traits in youths. Nat. Rev. Neurosci. *14*, 786–799.
3. Smith, A. (1759). Theory of Moral Sentiments (Cambridge).
4. Blair, R.J.R., Budhani, S., Colledge, E., and Scott, S. (2005). Deafness to fear in boys with psychopathic tendencies. J. Child Psychol. Psychiatry *46*, 327–336.

5. Meffert, H., Gazzola, V., den Boer, J.A., Bartels, A.A.J., and Keysers, C. (2013). Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. Brain *136*, 2550–2562.
6. Keysers, C., and Gazzola, V. (2014). Dissociating the ability and propensity for empathy. Trends Cogn. Sci. *18*, 163–166.
7. Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H.E., Lin, S.-Y.Y., Rabah, D., Kinet, J.-P.P., and Shin, H.-S.S. (2010). Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. Nat. Neurosci. *13*, 482–488.
8. Kim, E.J., Kim, E.S., Covey, E., and Kim, J.J. (2010). Social transmission of fear in rats: the role of 22-kHz ultrasonic distress vocalization. PLoS ONE *5*, e15077.
9. Atsak, P., Orre, M., Bakker, P., Cerliani, L., Roozendaal, B., Gazzola, V., Moita, M., and Keysers, C. (2011). Experience modulates vicarious freezing in rats: a model for empathy. PLoS ONE *6*, e21855.
10. Li, Z., Lu, Y.F., Li, C.L., Wang, Y., Sun, W., He, T., Chen, X.F., Wang, X.L., and Chen, J. (2014). Social interaction with a cagemate in pain facilitates subsequent spinal nociception via activation of the medial prefrontal cortex in rats. Pain *155*, 1253–1261.
11. Meyza, K.Z., Bartal, I.B., Monfils, M.H., Panksepp, J.B., and Knapska, E. (2017). The roots of empathy: through the lens of rodent models. Neurosci. Biobehav. Rev. *76* (Pt B), 216–234.
12. Burkett, J.P., Andari, E., Johnson, Z.V., Curry, D.C., de Waal, F.B., and Young, L.J. (2016). Oxytocin-dependent consolation behavior in rodents. Science *351*, 375–378.
13. de Waal, F.B.M., and Preston, S.D. (2017). Mammalian empathy: behavioural manifestations and neural basis. Nat. Rev. Neurosci. *18*, 498–509.
14. Ferretti, V., Maltese, F., Contarini, G., Nigro, M., Bonavia, A., Huang, H., Gigliucci, V., Morelli, G., Scheggia, D., Managò, F., et al. (2019). Oxytocin signaling in the central amygdala modulates emotion discrimination in mice. Curr. Biol. *29*, 1938–1953.e6.
15. Scheggia, D., Managò, F., Maltese, F., Bruni, S., Nigro, M., Dautan, D., Latuske, P., Contarini, G., Gomez-Gonzalo, M., Requie, L.M., et al. (2020). Somatostatin interneurons in the prefrontal cortex control affective state discrimination in mice. Nat. Neurosci. *23*, 47–60.
16. Han, Y., Bruls, R., Soyman, E., Thomas, R.M., Pentaraki, V., Jelinek, N., Heinemans, M., Bassez, I., Verschooren, S., Pruis, I., et al. (2019). Bidirectional cingulate-dependent danger information transfer across rats. PLoS Biol. Published online December 5, 2019. https://doi.org/10.1371/journal.pbio.3000524.
17. Bruchey, A.K., Jones, C.E., and Monfils, M.-H. (2010). Fear conditioning by-proxy: social transmission of fear during memory retrieval. Behav. Brain Res. *214*, 80–84.
18. Jones, C.E., Riha, P.D., Gore, A.C., and Monfils, M.-H. (2014). Social transmission of Pavlovian fear: fear-conditioning by-proxy in related female rats. Anim. Cogn. *17*, 827–834.
19. Sakaguchi, T., Iwasaki, S., Okada, M., Okamoto, K., and Ikegaya, Y. (2018). Ethanol facilitates socially evoked memory recall in mice by recruiting pain-sensitive anterior cingulate cortical neurons. Nat. Commun. *9*, 3526.
20. Carrillo, M., Han, Y., Migliorati, F., Liu, M., Gazzola, V., and Keysers, C. (2019). Emotional mirror neurons in the rat's anterior cingulate cortex. Curr. Biol. *29*, 1301–1312.e6.
21. Allsop, S.A., Wichmann, R., Mills, F., Burgos-Robles, A., Chang, C.J., Felix-Ortiz, A.C., Vienne, A., Beyeler, A., Izadmehr, E.M., Glober, G., et al. (2018). Corticoamygdala transfer of socially derived information gates observational learning. Cell *173*, 1329–1342.e18.
22. Cruz, A., Heinemans, M., Marquez, C., and Moita, M.A. (2019). Freezing displayed by others is a learned cue of danger resulting from co-experiencing own-freezing and shock. bioRxiv. https://doi.org/10.1101/800714.
23. Ben-Ami Bartal, I., Rodgers, D.A., Bernardez Sarria, M.S., Decety, J., and Mason, P. (2014). Pro-social behavior in rats is modulated by social experience. eLife *3*, e01385.
24. Pitcher, M.H., Gonzalez-Cano, R., Vincent, K., Lehmann, M., Cobos, E.J., Coderre, T.J., Baeyens, J.M., and Cervero, F. (2017). Mild social stress in mice produces opioid-mediated analgesia in visceral but not somatic pain states. J. Pain *18*, 716–725.
25. Pisansky, M.T., Hanson, L.R., Gottesman, I.I., and Gewirtz, J.C. (2017). Oxytocin enhances observational fear in mice. Nat. Commun. *8*, 2102.
26. Langford, D.J., Tuttle, A.H., Brown, K., Deschenes, S., Fischer, D.B., Mutso, A., Root, K.C., Sotocinal, S.G., Stern, M.A., Mogil, J.S., and Sternberg, W.F. (2010). Social approach to pain in laboratory mice. Soc. Neurosci. *5*, 163–170.
27. Martin, L.J., Hathaway, G., Isbester, K., Mirali, S., Acland, E.L., Niederstrasser, N., Slepian, P.M., Trost, Z., Bartz, J.A., Sapolsky, R.M., et al. (2015). Reducing social stress elicits emotional contagion of pain in mouse and human strangers. Curr. Biol. *25*, 326–332.
28. Langford, D.J., Crager, S.E., Shehzad, Z., Smith, S.B., Sotocinal, S.G., Levenstadt, J.S., Chanda, M.L., Levitin, D.J., and Mogil, J.S. (2006). Social modulation of pain as evidence for empathy in mice. Science *312*, 1967–1970.
29. Cui, F., Abdelgabar, A.R., Keysers, C., and Gazzola, V. (2015). Responsibility modulates pain-matrix activation elicited by the expressions of others in pain. Neuroimage *114*, 371–378.
30. Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage *54*, 2492–2502.
31. Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., and Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. Science *303*, 1157–1162.
32. Vogt, B.A., and Paxinos, G. (2014). Cytoarchitecture of mouse and rat cingulate cortex with human homologies. Brain Struct. Funct. *219*, 185–192.
33. Vogt, B.A.A., and Peters, A. (1981). Form and distribution of neurons in rat cingulate cortex: areas 32, 24, and 29. J. Comp. Neurol. *195*, 603–625.
34. Paxinos, G., and Watson, C. (1998). The Rat Brain in Stereotaxic Coordinates, Fourth Edition (Academic).
35. Paxinos, G., and Watson, C. (2013). The Rat Brain in Stereotaxic Coordinates, Seventh Edition (Elsevier).
36. Martin, J.H. (1991). Autoradiographic estimation of the extent of reversible inactivation produced by microinjection of lidocaine and muscimol in the rat. Neurosci. Lett. *127*, 160–164.
37. Greene, J.T. (1969). Altruistic behavior in the albino rat. Psychon. Sci. *14*, 47–48.
38. Hernandez-Lallement, J., van Wingerden, M., and Kalenscher, T. (2018). Towards an animal model of callousness. Neurosci. Biobehav. Rev. *91*, 121–129.
39. Han, Y., Sichterman, B., Carrillo, M., Gazzola, V., and Keysers, C. (2019). Similar levels of emotional contagion in male and female rats. bioRxiv. https://doi.org/10.1101/857094.
40. Langford, D.J., Tuttle, A.H., Briscoe, C., Harvey-Lewis, C., Baran, I., Gleeson, P., Fischer, D.B., Buonora, M., Sternberg, W.F., and Mogil, J.S. (2011). Varying perceived social threat modulates pain behavior in male mice. J. Pain *12*, 125–132.
41. Rogers-Carter, M.M., Djerdjaj, A., Culp, A.R., Elbaz, J.A., and Christianson, J.P. (2018). Familiarity modulates social approach toward stressed conspecifics in female rats. PLoS ONE *13*, e0200971.
42. Ishii, A., Kiyokawa, Y., Takeuchi, Y., and Mori, Y. (2016). Social buffering ameliorates conditioned fear responses in female rats. Horm. Behav. *81*, 53–58.
43. Gruene, T.M., Flick, K., Stefano, A., Shea, S.D., and Shansky, R.M. (2015). Sexually divergent expression of active and passive conditioned fear responses in rats. eLife *4*, e11352.
44. Knapska, E., Mikosz, M., Werka, T., and Maren, S. (2009). Social modulation of learning in rats. Learn. Mem. *17*, 35–42.

45. Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: An Introduction (MIT).
46. Hutchison, W.D., Davis, K.D., Lozano, A.M., Tasker, R.R., and Dostrovsky, J.O. (1999). Pain-related neurons in the human cingulate cortex. Nat. Neurosci. *2*, 403–405.
47. Han, Y., Bruls, R., Soyman, E., Thomas, R.M., Pentaraki, V., Jelinek, N., Heinemans, M., Bassez, I., Verschooren, S., Pruis, I., et al. (2019). Bidirectional cingulate-dependent danger information transfer across rats. PLoS Biol. *17*, e3000524.
48. Sanders, J., Mayford, M., and Jeste, D. (2013). Empathic fear responses in mice are triggered by recognition of a shared experience. PLoS ONE *8*, e74609.
49. Batson, C.D., O'Quin, K., Fultz, J., Vanderplas, M., and Isen, A.M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. J. Pers. Soc. Psychol. *45*, 706–718.
50. Barker, D.J., Sanabria, F., Lasswell, A., Thrailkill, E.A., Pawlak, A.P., and Killeen, P.R. (2010). Brief light as a practical aversive stimulus for the albino rat. Behav. Brain Res. *214*, 402–408.
51. Reed, P., and Yoshino, T. (2008). Effect of contingent auditory stimuli on concurrent schedule performance: an alternative punisher to electric shock. Behav. Processes *78*, 421–428.
52. Coffey, K.R., Marx, R.G., and Neumaier, J.F. (2019). DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. Neuropsychopharmacology *44*, 859–868.
53. Wöhr, M., and Schwarting, R.K.W. (2013). Affective communication in rodents: ultrasonic vocalizations as a tool for research on emotion and motivation. Cell Tissue Res. *354*, 81–97.
54. Jourdan, D., Ardid, D., Chapuy, E., Eschalier, A., and Le Bars, D. (1995). Audible and ultrasonic vocalization elicited by single electrical nociceptive stimuli to the tail in the rat. Pain *63*, 237–249.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Betadine	Daxtrio Medische Producten	Betadine 100 mg/ml
Dental Cement	Prestige Dental	Super Bond C&B Kit
Isoflurane	Veterinary Technics	IsoFlo
Lidocaine	AstraZeneca	Xylocaine 10% Pump Spray
Meloxicam	Boehringer Ingelheim	Metacam 5 mg/ml
Muscimol	Sigma Aldrich	M1523
Deposited Data		
Data	This paper	https://osf.io/65j3g/
Experimental Models: Organisms/Strains		
Sprague Dawley rats	Janvier Labs	RjHan:SD
Software and Algorithms		
Avisoft-RECORDER (version 4.2.24)	Avisoft Bioacoustics	https://www.avisoft.com/recorder/
DeepSqueak (version 2.6.1)	[52]	https://github.com/DrCoffey/DeepSqueak
IBM SPSS Statistics (version 25)	IBM	https://www.ibm.com/products/spss-statistics
JASP (version 0.10.2.0)	JASP Team	https://jasp-stats.org/
MATLAB (version R2017b)	Mathworks	https://www.mathworks.com/products/matlab.html
Media Recorder (version 4.0.542.1)	Noldus	https://www.noldus.com/human-behavior-research/products/mediarecorder
MED-PC IV (version 4)	Med Associates	https://www.med-associates.com/med-pc-v/
R (version 1.1.463)	R Foundation	https://www.r-project.org/
Solomon Coder (version beta 17.03.22)	András Péter	https://solomon.andraspeter.com/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Christian Keysers (c.keysers@nin.knaw.nl). This study did not generate any new reagents or animal lines.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects

A total of 314 Sprague Dawley rats were ordered from Janvier Labs (France). Rats were separated in different groups to test different parameters of the harm aversion paradigm (Tables 1 and 2). Except for one group specifically testing females (Table 1), all animals were males. Rats were socially housed in groups of four same-sex individuals (SPF, type III cages with sawdust), in a temperature-(22-24°C) and humidity-controlled (55% relative humidity) animal facility, on a reversed 12:12 light:dark cycle (lights off at 07:00). Water was provided *ad libitum*. Upon start of the training phase, a food deprivation schedule was implemented to maintain animals at 85% of their free-feeding body weight, which was monitored daily. Animals were pre-fed with 66% of their daily food intake 2h before the start of harm aversion testing (to prevent that high hunger masks harm aversion). All tests were done in the dark phase of the animals' circadian rhythm, between 08:30 and 13:00. All animals were 30 days old at arrival, and started harm aversion testing at 55 days of age. Male and female rats weighed on average 302.4g (SD = 96.5) and 240.8g (SD = 14.1) at the start of the experiment. All experimental procedures were approved by the Centrale Commissie Dierproeven of the Netherlands (AVD801002015105) and by the welfare body of the Netherlands institute for Neuroscience (IVD, protocol number NIN171105-181103). Studies were conducted in strict accordance with the European Community's Council Directive (86/609/EEC).

Sample size calculation

Behavioral experiments

One experiment performed in rats reported that changes in (n[rats] = 10) preferences contingent on caused conspecific distress was significant at p = 0.01 [37]. Based on z-score tables (http://www.z-table.com/), we estimated that an equivalent z-score = 2.29, which

outputted an effect size r = 0.72 ($r = zscore/(sqrt(N))$). For an effect size r = 0.72, an 80% power, $\alpha = 0.05$, we computed a required sample size of n = 8 actors per condition. We thus aimed for groups of at least 8 actors in all conditions, and sometimes used larger samples to increase sensitivity. We supplement the frequentist approach with Bayesian analysis whenever no significant *p-values* were found to further inform the interpretation of negative findings.

Pharmacology experiment

Given that the decrease in vicarious freezing in ACC deactivated animals corresponds to an effect size r = 1.22 [20], we used the same effect size for sample size calculations in the pharmacological experiment. For an effect size r = 1.22, an 80% power, $\alpha = 0.05$, we computed a required sample size of n = 12 animals. Accounting for ~10% of dropout (e.g., surgery casualties, wrong target of the implanted cannula), we expect n = 14 actors for each muscimol and saline condition, summing to a total of n = 28 animals.

METHOD DETAILS

Experimental setup

The experimental setup (Figures 1A–1D) consisted of two skinner boxes (L: 30.5 cm x W: 24.1 cm x H: 29.2 cm; ENV-008CT; Med Associates, Inc.) fused into one single setup in our Mechatronics department. The separation walls between the compartments (i.e., skinner boxes) were replaced by a single perforated Plexiglas wall that allowed the transmission of auditory, olfactory and visual information between compartments. The compartment's floor consisted of a stainless grid floor (ENV-005). One of the compartments' floor (the victim's compartment) was linked to a stimulus scrambler that allow the delivery of a foot-shock (ENV-414S). In the actor's compartment, one nose poke unit (ENV-114BM) was installed on the wall opposite to the divider (Figures 1B and 1C). Two retractable levers (ENV-112CM) equipped with a stimulus light above each of them (ENV-221M) and a food hopper (ENV-200R2MA) beside each of them, were placed on the lateral walls of the actor's compartment (Figures 1C and 1D). Levers were equidistant from the nose poke unit. Two food dispensers (ENV-203-45IR) placed outside on each side the actor's compartment allowed the delivery of sucrose reward pellets for correct lever presses in the food hoppers (Figure 1B). Finally, two house lights (ENV-215M) were placed close to the top of the box above each dispenser (Figures 1A and 1B). House lights were turned ON before and after the session, which indicated that no operant items could be operated by the animals. The setups were placed in sound attenuating cabinets (Drefa, the Netherlands; Figure 1A). Rats performed all tasks in the dark. All tasks were controlled by custom scripts written in MED-PC IV (Med Associates, Inc.). Infrared cameras (one per compartment; 600 TVL 6 mm; Sygonix) were used to record the rats' behavior using the Media Recorder 4.0.542.1 (Noldus Information Technology) software. Auditory signals were recorded via a single condenser ultrasound microphone (CM16/CMPA, Avisoft Bioacoustics) and an ultrasound recording system (Sampling rate: 250 kHz; UltraSoundGate 416H and RECORDER 4.2.24 software, Avisoft Bioacoustics). The microphone was located next to the perforated divider separating the actor and victim compartments (Figures 1A and 1B), and this location, as well as the gain setting on the recording system, were kept exactly the same across dyads.

Experimental procedures

Acclimation, handling, and habituation

On the day of arrival, rats were randomly housed in groups of four and were allowed to acclimate for four days to the colony room. The animals' tails were marked from 1 to 4 to identify the animals in each cage. Numbering was random to avoid that animals less anxious would be labeled first in each cage. After 4 days, rats were handled during the dark phase of the cycle for 5 minutes for five days. Individuals #1 & #3 of each cage (1 and 3 stripes) were assigned to the actor role and individuals # 2 & #4 (2 and 4 stripes) were assigned to the victim role. Within each cage, two pairs were formed always following the same pattern: ACT#1 and VIC#2 formed one pair and ACT#3 and VIC#4 the second. The two members of a pair were always tested together, thus ensuring familiarity. There was only one exception to this rule: the Unfamiliar Victims condition that specifically investigated the effect of unfamiliarity within a dyad. On the fifth day of handling, animals were transported within their home cage to the experimental room, where the rats were placed in their compartment for 5 minutes for habituation, then weighed and placed back in their home cage. Between each rat, the compartment of the skinner box was cleaned with 70% ethanol. This was done consistently throughout each phase of the whole experiment. To facilitate the acquisition of reward-driven lever press, sucrose pellets (n = 3) were placed in the reward hopper on the first day of training. From this day on, the animals' food intake was reduced to bring animals to 85% of their free feeding body weight, which was monitored on a daily basis.

Training

On the session following habituation to the setup, the actors were shaped to press levers to obtain food. To do so, animals went through 3 consecutive steps of training (1 session/day).

Step 1. Actors were placed in their compartment, with house light on. The session was started by the experimenter using an adjacent computer, which was indicated to the animal by turning off the house light. Both levers were constantly presented and could be pressed by the animal. Either lever required 30cN to be operated. No time out was used to maximize the chance for animals to accidentally press one of the levers. Either lever press led to the delivery of n = 3 sucrose pellets in the adjacent reward hopper. Animals had a maximum session length of 20min, and could perform a maximum of 50 lever presses. Hence, animals could perform the session in less than 20min if they performed all 50 presses. The session was repeated on another day until the rat performed at least 35 lever presses out of the possible 50 (70%) within the 20min, in which case they were promoted to Step

2. This step was performed on average within four sessions (mean computed over ContingentHarm, NoHarm and RandomHarm = 3.78 sessions, SD = 1.41).

Step 2. Actors went through 5 blocks of 10 trials, leading to a total of 50 trials per session. Each block started with 4 forced trials (only one lever extended, 2 trials per lever, randomized order) followed by 6 free choice trials (both levers extended). Each trial started with the extension of the lever(s), and actors had 20 s to press one of the two extended levers. Animals could press only one lever per trial, i.e., pressing one lever led to the retraction of the opposite lever and the delivery of reward in the adjacent receptacle. Either levers required 30cN to be operated. Lever press led to the delivery of n = 1 sucrose pellet as well as the activation of the light cue located above the lever for 1 s. Pellets were delivered 1 s after lever press. Animals had a maximum of 30min to perform all 50 trials. The 4 forced trials of each block were mandatory, and timeout (20 s without pressing) led to the repetition of that trial. For free choice trials, a timeout led to the retraction of the levers and switching to the next trials. Promotion criterion to Step 3 was to press any lever within 20 s on at least 21 of the 30 free choice trials (70%) within a session. This step was performed on average within two sessions (mean computed over ContingentHarm, NoHarm and RandomHarm = 2.22 sessions, SD = 0.51).

Step 3. Here, a nose poke was required in order to activate the levers and initiate the trial. In the first session, the nose-poke light was turned on for 50 s, and a 10ms nose poke within that 50 s was enough to trigger the extension of the levers. Animals then had 20 s to press a lever, and perform a trial correctly. Performing at least 70% correct trials of the 30 free choice trials was used as a criterion to increase the required nose poke duration to 200ms, and the duration of the nose-poke illumination to 20 s. Again, if 70% of the free choice trials are performed correctly, animals arrived at the final nose poke duration of 400ms to ensure that animals initiated the trial deliberately and not by accident. A timeout (i.e., maximum time window for response) of 20 s was implemented for nose poke and lever press, which ensured that animals performed trials in a rapid fashion. Hence, animals performed a minimum of 150 trials (i.e., three levels on nose poke duration) in this step. For the conditions testing magnitude (1vs2Pellets and 1vs3Pellets) and delay (0vs2s), the levers required 30cN to be operated and led to one pellet each during training. For all other conditions, one of the levers was modified so as to double the number of newtons required to operate it (60cN versus 30cN). This adjustment was done at the beginning of every session using a dynamometer (DM10, Nouvoutils). The hard lever's side was randomized between rats, but was stable within rats. Animals had a maximum of 30min to perform all 50 trials. Animals were considered ready for testing when rats performed at least 70% of the free choice trials within the final parameters. Since the nose poke duration had three increasing duration steps, this last step of training consisted of a minimum of three sessions of 50 trials each. This step was performed on average within four sessions (mean computed over ContingentHarm, NoHarm and RandomHarm = 3.85 sessions, SD = 0.67).

Habituation actor & victim

When actors had completed the training phase, the dyads were placed in the testing setup for 10min with house lights on, in order to habituate (i) the victim to the novel environment and (ii) the actor to the victim's presence.

Exposure

When the actors were habituated to the presence of their partner, they were placed in the victim's compartment (''Exposure''; Figure 1F) and were exposed to electric shocks [9]. This procedure consisted of a baseline (10min) followed by the delivery of four foot-shocks (0.8mA, 1 s duration, 240-360 s random inter-shock interval; yellow lightening in Figure 1). Actors were exposed alone in the setup. After the session, the actors were isolated in a small cage for 10 minutes to allow for the following animals to be placed in the exposure setup, so as to avoid the communication of stress to unexposed cage mates. For actors in the NonExposed condition (see below, and Table 2), the shockers were turned off during the Exposure session. For one group of rats, (Unfamiliar Victims condition), the exposure was performed in other cabinets using different odors and wall design [9], for historical reasons, as this condition was part of earlier experiments aimed to fine-tune our final experimental design.

Harm aversion testing

All dyads underwent four harm aversion testing sessions. The first session (''Baseline''; Figures 1E and 1F) took place on the day following the exposure, and was followed by 3 consecutive daily ''Shock'' sessions. Baseline and Shocks sessions were identical in all points except that during Shock sessions, pressing one of the two levers immediately delivered an electric foot-shock (0.8mA, 1 s duration) to the victim (Figure 1F, ''Shock''), while shocks were never delivered during baseline (Figure 1F, ''Baseline''). Hence, the baseline sessions allowed to sample initial actor preferences for a given lever. While most actors preferred the easy lever in the baseline session (~75%), a subset preferred the hard lever. The lever that delivered a shock to the victim during harm aversion testing was determined based on preference levels at baseline. This did not affect group differences in harm aversion behavior (Figure S4A). However additional training prior to harm aversion testing competed with harm aversion (Figure S4B). Baseline and Shock sessions started with 4 forced trials (2 for each lever, pseudo-randomized) followed by 20 free choice trials (Figure 1E). This structure was constant for all conditions, except for the actors of the Unfamiliar Victims condition, where two baseline sessions followed by two shock sessions (each consisting of 4 forced and 8 free choice trials) were used for historical reasons. The forced trials were implemented to force the animals to sample each lever-outcome contingency at least twice per session. Failure to perform the forced trials restarted the trial (rats needed to perform all forced trials, i.e., were forced to sample each option twice). Failure to perform a free choice trial resulted in a failed trial, which incremented to the following trial. Animals had to respond within a time window of 400 s otherwise the current trial was counted as failed (Figure 1G). The trial time-out for the Baseline, Shock and Food sessions was increased to 400 s to ensure that even if actors were to have heightened attention toward their victim in distress, they would have enough time to initiate a novel trial. Animals were pre-fed 2h before beginning of each session with 66% of their daily food intake

to avoid high levels of hunger which could have competed with harm aversion. The remaining 34% of the food allowance was provided after the experiment.

Experimental conditions

Several different experimental conditions, each of which used a separate group of animals, were tested (Tables 1 and 2). Each actor was paired with a unique victim (sample size in Tables 1 and 2 describes only the number of actors). Except for the familiarity control (see "Unfamiliar Victims" condition), actors and victims were housed in the same cage and were thus familiar with one another for 20 days before harm aversion testing started. In all cases animals were always tested with sex-matched conspecifics. Except for half of the dyads in the ContingentHarm condition (n = 12 actors), all tested dyads were males.

The effect of shock contingency. Rats were assigned to three main conditions: ContingentHarm (n = 12 male, n = 12 female), NoHarm (n = 14), or RandomHarm (n = 8) conditions. The ContingentHarm condition tested whether rats would decrease their preference for a lever if pressing that lever triggered a conspecific's distress. In this condition, the victim was immediately administered an electric shock every time the actor pressed the shock lever in the shock sessions, while rewards were administered 1 s after lever press. Shock-lever association was deterministic and the shock intensity was identical to the one used during the exposure session of the actor (1 s duration, 0.8mA). In order to control for spontaneous changes in lever pressing (i.e., not related to the shocks to conspecific), victims in the NoHarm condition were constantly isolated from shocks by using a plastic isolating floor during shock sessions. Hence, shocks were physically delivered to the grid floor, but never reached the victim. This was done to ensure that sounds associated with shock-delivery would be present in both conditions. With the RandomHarm condition, we aimed at disrupting the contingency between lever-presses and shocks, while keeping the level of victim distress comparable to that of the ContingentHarm. This condition therefore served as a control to test whether animals in the ContingentHarm condition that switched to the no-shock lever did so because their victim was stressed, or because the distress of the victim was specifically associated with pressing the shock lever. We thus selected the 8 pairs in the ContingentHarm condition that showed the strongest switch away from the shock lever, and yoked each one of them to a pair in the RandomHarm condition. The 8 victims used in the RandomHarm condition then received the same amount of shocks, and at the same trial number, as the 8 highest switcher animals in the ContingentHarm condition (determined using a permutation test; see "Statistical use of indexes"). Hence, victim stress levels in RandomHarm and ContingentHarm victims were comparable (as verified in Figure S1). Crucially however, (i) the shocks were not delivered at lever press but at random intervals ranging between 3 s and 8 s during the inter-trial interval, which started after the actor exited the reward receptacle and (ii) shocks were delivered independently of what lever was pressed by the actor. This ensured that the contingency between lever-pressing and victim distress was perturbed, while total distress was kept similar.

The effect of familiarity. To explore the effect of familiarity between actors and victims on harm aversion, an additional group of rats (n = 12) was piloted with the above-mentioned procedures. Critically, the actors in this Unfamiliar Victims condition were paired with victims housed in a different cage, thus ensuring that the rats within each dyad were unfamiliar with each other.

The effect of type and degree of cost to self. Additional groups were used to test the effects of different types and degrees of cost to the actor on harm aversion. These groups were treated the same as described for the ContingentHarm condition, except the following differences. The magnitude of cost was manipulated in 1vs2Pellets (n = 7) and 1vs3Pellets conditions (n = 11). In these conditions, the levers required the same amount of force to be operated, however the shock lever delivered 2 (1vs2Pellets) and 3 (1vs3Pellets) pellets, while the no-shock lever delivered 1 pellet. In these conditions, actors were trained with levers of equal effort (30cN required to operate the levers) and equal reward contingencies (1 pellet per lever during training). After training, animals went through additional sessions (M = 2.72 sessions, SD = 0.91 in 1vs3Pellets group; M = 3.25 sessions, SD = 0.46 in the 1vs2Pellets group) where one lever was associated with three (1vs3Pellets) or two pellets (1vs2Pellets). These additional training sessions were implemented to habituate the rats to the novel reward contingencies before collecting the baseline data for the experiment.

The effect of shock experience: To assess whether actor's prior experience with footshocks modulate harm aversion, actors in the NonExposed condition (n = 14) underwent the same treatment as the ContingentHarm animals, except that they received no shock during the exposure session.

The effect of ACC deactivation: Finally, the effect of deactivating the ACC on harm aversion was tested by using two additional groups of rats (Muscimol, n = 11; Saline, n = 12). These rats underwent identical procedures as the ContingentHarm condition, but were bilaterally implanted with internal cannulas targeting the ACC to inject muscimol or saline, respectively (see below).

Food control

A lack of behavioral flexibility could account for actors not switching away from the shock lever. To test this possibility, all actors (except for 1vs2Pellets, 1vs3Pellets, 0vs2s, and Unfamiliar Victims groups) underwent 3 consecutive food sessions (1 session/day; Figures 1E and 1F; Tables 1 and 2) upon completion of the shock sessions. In food sessions, the lever which was not preferred (pressed < 50% of the trials) during the last Shock session was associated with a higher reward (3 pellets) while the other lever still delivered the same amount (1 pellet). The actors went through three sessions of 24 trials (identical to the baseline session structure; Figure 1E) with the victim present in the adjacent compartment but without shocks involved during the session (Figure 1F, "Food"). Hence, the context was similar to the baseline and shock sessions, but the animal could reverse their previously acquired preferences to obtain more food.

Trial structure

Trial structure was identical across Baseline, Shock, and Food sessions (Figure 1G). The trial started with illumination of the nose poke hole, inviting a nose poke. Actors were expected to perform a 400ms nose poke which triggered the presentation of the levers (only one lever in forced trials, both levers on free trials). Pressing either lever led to the delivery of sucrose pellets after 1 s in adjacent

reward receptacles (n = 1 pellet for each lever in baseline and shock sessions for all conditions except 1v2Pellets and 1v3Pellets, in which one lever led to 1 pellet and the other to 2 or 3 pellets, respectively; n = 1 and n = 3 pellets during food sessions), with the hopper illuminating for 0.5 s upon delivery of the reward. During shock sessions (except for the NoHarm and RandomHarm conditions) pressing the shock lever additionally led to an immediate 1 s shock to the victim. An inter-trial interval of 10 s started once the animal had exited the reward hole to consume the pellets, after which the nose poke was illuminated for the next trial to start. During baseline and shock sessions, a maximum response time window of 400 s was used to stop the current trial in case the animal did not perform a nose poke or did not press levers after the nose poke (Figure 1G; "Time Out"). This was deemed long enough to show a behavioral display of harm aversion (longer latency to consume food), while quantifying failed trials. In each trial, we collected two timing measurements happening after lever press: reward poke latency (from successful lever press to first entry in reward receptacle) and reward poke duration (time spent consuming the reward). Forced trials were systematically excluded from all analysis since (i) they did not reflect actual decisions from the actors and (ii) latencies might have been affected by acclimation time to the session.

Surgery and cannulation

After training was completed, rats in the muscimol and saline conditions (N_{total} = 27) underwent a surgical procedure for the bilateral implantation of cannulas targeting the ACC. Body temperature and other physiological parameters were monitored throughout the surgery. Rats were anaesthetized using isoflurane (5% induction, 2/2.5% maintenance), and prepared for surgery by weighing and shaving the head. Once animals were placed on a stereotaxic apparatus, the incision area was cleaned with alcohol/betadine and sprayed with 10% xylocaine (lidocaine, spray) used as a local anesthetic. Two holes were drilled to bilaterally implant with stainless steel cannulas targeting the ACC (Plastic One, C313G/Spc 3.5mm). The cannulaes were placed with a 20° angle from the skull's surface at the following coordinates: AP, + 1.17 mm; ML, μ1.16 mm; DV, +1.8 mm (AP & ML from bregma; DV from the surface of the skull [20];). All coordinates were taken based on [34]. Three additional holes were drilled around the implanted cannula to place steel screws, which were later used to anchor the cannula to the skull using dental cement (Prestige Dental, Super Bond C&B Kit, UK). To minimize the damage to the cannula, they were covered with a protection cap (Plastic One, C313DC/1/Spc 3.5mm). After the surgery, an analgesic/anti-inflammatory drug was delivered for pain relief (meloxicam, 2 mg/kg, sc) and 0.5ml of saline sc was given for rehydration. Animals were then placed in an incubator until they woke up. The animals were housed individually for 10 days to recover from the surgery. To monitor any possibility of discomfort or pain and to ensure that the animals were having a proper recovery process, the appearance, behavior, state of the incision (wound healing), recovery process, and weight were monitored daily for 10 days after the surgery. Two animals died during surgery. Post-mortem dissection revealed a small heart lumen which might have rendered animals more sensitive to isofluorane anesthesia. One additional animal showed abnormal weight loss and aberrant behavior, and was euthanized two days after surgery (euthanized in CO_2 chambers with initial 40% O_2 mixed with 60% CO_2 until animals were in deep sleep as verified through paw reflexes, then switched to 100% CO2 for at least 15 minutes). For the remaining 24 animals, following the surgery, all rats had minimal or no weight loss, and the animals that lost some weight returned to normal weight following 2 to 3 days after the surgery. In addition, all these animals showed normal behavior, prompt recovery, and healthy wound healing following the surgery. After 10 days of recovery, animals were tested two days in the operant box (same design as training, step 3) to ensure that the implanted cannula did not affect food access, and that operant behavior was unaffected by the intervention.

Humane endpoints

The humane endpoints were as follows

Insufficient recovery after surgery. It was considered if animal showed permanent weight loss. The threshold was set to a 15% weight loss after surgery monitored during 10 days. One animal was euthanatized following this criterion.

Infection. Although we always perform the surgeries in sterile conditions, there was a small possibility of infection around the wound area. Visible signs of pathogenesis were monitored. The following were considered as signs of unhealthy state of the animal: aberrant behavior, shock, dehydration, weight loss, nose and mouth discharge, bleeding, fits/seizures, and diarrhea. No animal was euthanized following this criterion.

Infusion of saline and muscimol

Actor rats in the Muscimol condition were given micro-injections of the GABAa receptor agonist muscimol before the start of the behavioral testing in baseline, shock and food sessions. Muscimol (Sigma Aldrich, M1523) was dissolved in sterile phosphate buffered saline to obtain a final concentration of 0.1ug/ul. Infusions (0.5ul per hemisphere) were made bilaterally under isoflurane anesthesia, which lasted on average 7min. This time included 2min of induction, 2.5min of infusion time (0.25 ul/min infusion rate), and 2.5min of diffusion time (with infusion needle left in place). Infusions were made in both hemispheres simultaneously using an infusion pump (Syringe Pump PHD Ultra Infuse, Harvard Apparatus) equipped with a micro-dialysis rack for 4 syringes (Harvard Apparatus). In the Saline condition, the exact same procedure was followed except that phosphate buffered saline (0.9%) was used without muscimol. From the end of infusion, a delay of 20min was implemented before the start of the sessions to allow full effect of muscimol on brain tissue.

Histology

Upon completion of the food sessions, animals were anaesthetized using isoflurane (5% induction). Depth of anesthesia was checked by verifying the paw reflexes, after which animals were perfused transcardially using 0.01M phosphate buffered saline

(PBS, 0.1M, pH = 7.4) for 3min followed by a fixating solution of paraformaldehyde (PFA, 4%) for 5min. Brains were immediately removed and stored in PFA solution for 10 days at a temperature of 5°C. Coronal sections (50 um) of the ACC were obtained using a vibratome (Leica VT1000S, Germany) and mounted for histological examination. Finally, injection sites were mapped using a microscope (Zeiss Axioplan 2, Germany) and the rat atlas [35] with standardized coordinates. Due to methodological issues, during brain extraction and slicing, this analysis resulted in n = 9 brains mounted in each group (muscimol and saline). Coordinates shown in Figure 4 were the result of combining the evidence about location from both hemispheres. Victims were euthanized in CO_2 chambers with initial 40% O_2 mixed with 60% CO_2 until animals were in deep sleep, as verified through paw reflexes, then switched to 100% CO2 for at least 15min. One animal was discarded due to clogged internal cannula during infusion in the baseline session. The remaining animals (muscimol, n = 11; saline, n = 12) were used for behavioral analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of lever presses

In order to compare the effect of different manipulations at the group level, we used parametric frequentist (IBM SPSS Statistics 25, IBM) and Bayesian statistics (JASP 0.10.2.0, https://jasp-stats.org/). We did so, because over the groups, lever preferences were approximately normally distributed, as assessed using the Shapiro-Wilk test.

In order to explore how many individuals showed significant switching during the shock or food sessions, despite differences in baseline lever preferences, we also computed two indexes: the switching index, capturing changes in preference from baseline to shock sessions, and the food index, capturing changes in preference from shock to food sessions.

Switching index

In order to quantify individual levels of switching despite differences in baseline preference, we computed a Switching Index (SI) for each rat, using the following equation,

$$SI = \frac{S_{baseline} - S_{shock}}{S_{baseline} + S_{shock}}$$

where $S_{baseline}$ is the proportion of shock lever presses during baseline, and S_{shock} the average proportion of shock lever presses over all three shock sessions. The SIs are individual values that range between [-1/3;1] and quantify the strength of switching between baseline and shock sessions. The distribution of potential SIs can be found in Figure 1F. Possible values for $S_{baseline}$ were sampled uniformly from the interval [0.5;1], given that $S_{baseline}$ cannot be below 0.5, because the shock lever is the lever preferred at baseline, by definition. Possible values for S_{shock} were sampled from the interval [0;1].

Food index

In order to quantify the strength of food learning, we computed a Food Index (RI) for each individual. In food sessions, the least preferred lever during shock sessions was now associated with n = 3 pellets, whereas the remaining lever produced n = 1 pellets. The FI was computed using the following equation,

$$FI = \frac{L_{food} - L_{shock3}}{1 - L_{shock3}}$$

where L_{food} is the proportion of choice for the lever producing 3 pellets averaged over all food sessions, and L_{shock3} that during the last shock session. This normalizes the change in preference by the maximum potential change possible in preference. The FIs are individual values that range between -1 and 1 and quantify the preference for the 3 pellets option. The distribution of potential FIs can be found in Figure 1F. Possible values L_{shock3} were sampled uniformly from [0;0.5], because the 3 pellet lever is by definition that non-preferred in shock3. Possible values for L_{food} from [0;1].

Statistical analysis

The equations used for SI and FI create skewed distributions (Figure 1F). To overcome these issues, we used non-parametric permutation analysis to detect animals that showed significant switching. We generated, for each animal separately, a distribution of permuted indexes, computed by shuffling the choices (with replacement, 10.000 times) across baseline and shock sessions (SI), or between shock3 sessions and food sessions (FI). We then compared actual indexes to the 95% confidence interval (CI) of this randomized distribution of social bias scores. We provide the individual CI's lower and upper bound in the Supplemental Information (Tables S1, S2, and S3). In addition, to explore whether the observed effects could be due to the bias introduced by selecting the shock and food levers as those above and below 50%, respectively, we also concentrate analyses on comparing choices across experimental and control conditions, which suffer from the same bias.

Additional behavioral analysis

Video scoring

To get a more detailed view of the rats' behavior, the recorded videos were manually scored with the use of the program Solomon Coder beta 17.03.22 (https://solomon.andraspeter.com/). Behavioral analysis focused on variables that could assess distress,

attention. A first set of variables is meant to reflect distress (freezing of actor and victim as well as pain squeaks of victim and USV emissions, see audio analysis below) or attention (time spent close to the divider). Because most of the switching occurred in the first shock session, we quantify these variables in this first session, or as changes between the first session and baseline.

Statistical analysis

MedAssociates and Solomon behavioral logfiles data extraction was performed with the use of MATLAB R2017b (MathWorks). Further statistical analyses were performed with IBM SPSS Statistics 25 (IBM). Data are expressed as means μ standard errors of the means (sem). The significance level was set at $p < 0.05$. Unless specified otherwise, all the displayed results correspond to behavioral data of the actor. The choice preference for the shock lever over sessions was analyzed with a repeated-measures analysis of variance (*rm*ANOVA) with session as a within-subjects factor (baseline and shock 1-3). The above stated analysis was also used for food session data and freezing behavior. Post hoc pairwise comparisons are reported in the figures. Pairwise comparisons consisted of paired sample t tests when comparing different epochs of the same group, and independent sample t tests when comparing different groups within an epoch. Given the odd distribution of the food index and SI, we used the non-parametric equivalent (Wilcoxon tests, Mann-Whitney U test). Comparisons were adjusted for multiple comparisons using the Benjamini & Hochberg correction (i.e., false discovery rate) using the R statistical package 1.1.463 (using the command p.adjust(unc*p*, method = ''BH,'' n = length(unc*p*); https://www.r-project.org/), where unc*p* is the list of uncorrected p values). Latencies for lever press, reward poke, nose poke and reward poke duration were log-transformed to meet parametric testing assumptions.

Interpretation of Bayes factor

We computed Bayesian statistics (JASP 0.10.2.0, https://jasp-stats.org/) in order to provide additional insights into our effects, and help differentiate evidence of absence from absence of evidence. We always used the default priors in JASP. For t tests, BF_{-0} is the Bayes Factor reflecting the plausibility of our data under the hypothesized reduction (hence – in the index after BF) divided by that our our data under a hypothesis of no effect (hence 0). BF_{10} is the Bayes Factor reflecting the plausibility of the data under a hypothesis any change (be it increase or decrease, hence 1 for a two-tailed H_1) divided by that under a null hypothesis of no change. Conventionally, If $BF > 3$, there is moderate evidence in favor of the hypothesis in the nominator (i.e., H_-: reduction or H_1: change). If $BF_{10} < 1/3$, there is moderate evidence for H_0 (evidence for the absence of effect). If $1/3 < BF_{10} < 1$, no strong conclusion should be drawn based on this data regarding the existence or absence of effect, but the relative plausibility of both hypotheses can be appreciated. For ANOVAs, BF_{incl} is the Bayes Factor comparing the plausibility of the data under a model including a given effect divided by that under a model excluding a given effect, where effect can be the main effect or an interaction, and the magnitude of the BF_{incl} can be interpreted much like the BF10.

Audio analysis

22 kHz USV analysis

Rat ultrasonic vocalizations (USV) with frequencies around 22 kHz indicate negative affect [53]. To assess their effect on actors' harm aversion, audio recordings were first processed with DeepSqueak 2.6.1 toolbox in MATLAB [52] with the built-in Long Rat Call Network_V2 and all other default 'Detect Call' settings. This resulted in detection of all the lever presentation and lever press sounds, in addition to the 22 kHz and higher frequency USVs. A preliminary exploratory analysis revealed that 22 kHz USVs can be separated from all the other detected sounds by using 19-30 kHz principal frequency and > 0.4 tonality cutoffs. These parameters were verified via the automatic classification of an independent sample of randomly selected 10 audio recordings, which demonstrated high concordance with manual classifications (mean ± SD = 98 ± 2%, min = 92%, max = 100%). Thus, all original DeepSqueak detections were processed with these parameters to automatically extract 22 kHz USVs. The call length of these USVs were then used to estimate the proportion of total session time that was spent emitting 22 kHz USVs.

Squeak analysis

Rats emit pain squeaks (a.k.a. peeps) upon receiving shocks [54]. The loudness of the squeaks of victims during harm aversion testing sessions were quantified to investigate their impact on actor's harm aversion. Exploratory analyses revealed that squeaks started as soon as the actor pressed the shock lever and continued for approximately 2 s. The sounds produced by the lever press and release co-occurred with squeaks within the first 0.5 s. Thus, for each trial, the period starting 0.5 s after the lever press and continuing for 1.5 s was used to quantify the loudness of squeaks. The audio recording during this period was transformed to the frequency domain and the average power up to 18 kHz was calculated as the measure of squeak loudness. The 18 kHz frequency cutoff ensured that 22 kHz USVs were not included in the power estimation.

DATA AND CODE AVAILABILITY

All data used to generate the figures can be downloaded at https://osf.io/65j3g/.