Ovid Medline 341 results on 04/07/20 20 results on 11/03/20 with limit to dt=20200407-20201103	exp "Freunds Adjuvant"/ or (Freund* adj2 adjuvant).mp. or freund adjuvans.mp. OR (CFA adj3 inject*).mp. AND (Exp Anxiety/ OR Exp depression/ OR Exp avoidance learning/ OR Exp mood disorders/ OR Exp emotions/ OR Exp dark adaptation/ OR emotion*.mp. OR "negative affect".mp. OR anxiety*.mp. OR anxiogenic.mp. OR anxiolytic.mp. OR depression.mp. OR depressive.mp. OR "pro-depressant".mp. OR depressogenic.mp. OR antidepressant.mp. OR "exploratory behavior".mp. OR "social interaction".mp. OR "novelty suppressed feeding".mp. OR "novelty-induced hypophagia".mp. OR "sucrose preference".mp. OR "two bottle choice".mp. OR (light adj5 dark).mp. OR "marble burying".mp. OR "hole board".mp. OR "y-maze".mp. OR "elevated plus maze".mp. OR "elevated zero maze".mp. OR "elevated 0-maze".mp. OR "forced swim".mp. OR "tail suspension".mp. OR affective.mp. OR ((cage OR test*) adj3 (behavior* OR behaviour*)).mp.) AND (Exp rodentia/ OR rodent*.mp. OR rats.mp. OR rat.mp. OR mice.mp. OR mouse.mp. OR murid*.mp. OR murine.mp. OR murinae.mp. OR maze*.mp. OR paw*.mp.)
Embase 433 results on 04/07/20 40 results on 11/03/20 with limit [3-4-2020]/sd NOT [4-11-2020]/sd	('freund adjuvant'/exp OR ((freund* NEAR/2 adjuvant):ti,ab,kw,de) OR 'freund adjuvans':ti,ab,kw,de OR ((cfa NEAR/3 (inflammation OR inject*)):ti,ab,kw,de)) AND ('anxiety'/exp OR 'depression'/exp OR 'avoidance behavior'/exp OR 'mood disorder'/exp OR 'emotion'/exp OR 'dark adaptation'/exp OR 'elevated plus maze test'/exp OR emotion*:ti,ab,kw,de OR 'negative affect':ti,ab,kw,de OR anxiety*:ti,ab,kw,de OR anxiogenic:ti,ab,kw,de OR anxiolytic:ti,ab,kw,de OR depression:ti,ab,kw,de OR depressive:ti,ab,kw,de OR 'pro-depressant':ti,ab,kw,de OR depressogenic:ti,ab,kw,de OR antidepressant:ti,ab,kw,de OR 'exploratory behavior':ti,ab,kw,de OR 'social interaction':ti,ab,kw,de OR 'novelty suppressed feeding':ti,ab,kw,de OR 'novelty-induced hypophagia':ti,ab,kw,de OR 'sucrose preference':ti,ab,kw,de OR 'two bottle choice':ti,ab,kw,de OR ((light NEAR/5 dark):ti,ab,kw,de) OR 'marble burying':ti,ab,kw,de OR 'hole board':ti,ab,kw,de OR 'y-maze':ti,ab,kw,de OR 'elevated plus maze':ti,ab,kw,de OR 'elevated zero maze':ti,ab,kw,de OR 'elevated 0-maze':ti,ab,kw,de OR 'forced swim':ti,ab,kw,de OR 'tail suspension':ti,ab,kw,de OR affective:ti,ab,kw,de OR (((cage OR test*) NEAR/3 (behavior* OR behaviour*)):ti,ab,kw,de)) AND ('rodent'/exp OR rodent*:ti,ab,kw,de OR rats:ti,ab,kw,de OR rat:ti,ab,kw,de OR mice:ti,ab,kw,de OR mouse:ti,ab,kw,de OR murid*:ti,ab,kw,de OR murine:ti,ab,kw,de OR murinae:ti,ab,kw,de OR maze*:ti,ab,kw,de OR paw*:ti,ab,kw,de)
Web of Science 290 results on 04/07/20 29 results on 11/03/20 - Refined by: PUBLICATION YEARS: (2020)	1. TS=((Freund* near/2 adjuvant) or 'freund adjuvans' OR (CFA near/3 (inflammation OR inject*))) **2.** TS=(emotion* OR 'negative affect' OR anxiety* OR anxiogenic OR anxiolytic OR depression OR depressive OR 'pro-depressant' OR depressogenic OR antidepressant OR 'exploratory behavior' OR 'social interaction' OR 'novelty suppressed feeding' OR 'novelty-induced hypophagia' OR 'sucrose preference' OR 'two bottle choice' OR (light near/5 dark) OR 'marble burying' OR 'hole board' OR 'y-maze' OR 'elevated plus maze' OR 'elevated zero maze' OR 'elevated 0-maze' OR 'forced swim' OR 'tail suspension' OR affective OR ((cage OR test*) near/3 (behavior* OR behaviour*))) **3.** TS=(rodent* OR rats OR rat OR mice OR mouse OR murid* OR murine OR murinae OR maze* OR paw*) 4. 1 AND 2 AND 3
Scopus 421 results on 04/07/20 40 results on 11/03/20 with the following limit: (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020))	(TITLE-ABS-KEY ((Freund* w/2 adjuvant) or "freund adjuvans" OR (CFA w/3 (inflammation OR inject*)))) AND (TITLE-ABS-KEY (emotion* OR "negative affect" OR anxiety* OR anxiogenic OR anxiolytic OR depression OR depressive OR "pro-depressant" OR depressogenic OR antidepressant OR "exploratory behavior" OR "social interaction" OR "novelty suppressed feeding" OR "novelty-induced hypophagia" OR "sucrose preference" OR "two bottle choice" OR (light w/5 dark) OR "marble burying" OR "hole board" OR "y-maze" OR "elevated plus maze" OR "elevated zero maze" OR "elevated 0-maze" OR "forced swim" OR "tail suspension" OR affective OR ((cage OR test*) w/3 (behavior* OR behaviour*)))) AND (TITLE-ABS-KEY (rodent* OR rats OR rat OR mice OR mouse OR murid* OR murine OR murinae OR maze* OR paw*))

Table S1. Search terms for Ovid Medline, Embase, Web of Science, and Scopus.

OUTCOME ASSESSMENT ANALYSIS High: unblinded manual scoring reported Some concerns: automated analysis *or* blinded manual scoring reported Low: automated analysis *and* blinded manual confirmation reported Unspecified: method of analysis not reported
SELECTIVE OUTCOME REPORTING High: no open field locomotor activity reported for animals used in behavioral tests, incomplete reports for tests such as entries into open arms for EPM/EZM or into both LDB compartments Some concerns: alternatives to open field locomotor activity reported, such as wheel running Low: open field locomotor activity for animals used in behavioral tests is reported, entries into open arms for EPM/EZM or into light and dark compartments in LDB are reported
EXCLUSIONS High: reasons seem likely to introduce bias such as behavior related to tested outcome Some concerns: reasons for exclusions appear justified but have no standard threshold Low: explicitly report no exclusions, reasons for missing data unrelated to outcome behavior, exclusions balanced across groups Unspecified: no mention of exclusions or confirmed lack thereof
BASELINE CHARACTERISTICS High: no validation of CFA effect reported, no age/sex/weight-matching in population Low: von Frey or Hargreaves conducted on the same mice as the behavioral tests Unclear: unclear if von Frey, Hargreaves, or other conducted on the same mice as other tests
BLINDED EXPERIMENTER High: experimenter was reported not blinded to variables other than saline/CFA such as morphine treatment Low: experimenter was reported blinded to variables other than saline/CFA, such as to morphine treatment Unspecified: no report of blinding strategy or lack thereof NA: studies with saline vs. CFA injection as the only variable used in relevant behavioral tests were excluded on this criterion
CONFLICT OF INTEREST REPORT High: corporate funding or consulting fees reported Low: declared no competing or conflicting interests Unspecified: no report included in the study
INJECTION RANDOMIZATION Some concerns: reported as randomized but not mentioned how Unspecified: no report of randomization or confirmed lack thereof
PRIVATE FUNDING High: corporate funding or consulting fees reported Some concerns: funding source unclear Low: all grants for funding Unspecified: no funding sources reported
RANDOM SELECTION FOR OUTCOME ASSESSMENT Some concerns: report randomized but now how Unspecified: no report of randomization or confirmed lack thereof
SAMPLE SIZE CALCULATION Some concerns: sample size reported as considered, but unspecified how Low: sample size calculated, power analysis conducted Unspecified: no mention of how sample sizes were decided upon
WELFARE REGULATION COMPLIANCE Low: report compliance to animal welfare standards, approval by ethics committee

Table S2. Assessment rubric for risks of bias and quality assessments.

Note:
Data presentation: forest plots. Effect sizes are plotted as circles with 95% confidence intervals marked by the underlying line. Size of circle represents its weight in the random-effects summary. Color of circle matches a symmetrical scale extending from the lowest effect size in darkest purple to the highest effect size in green, with zero as white. Color scale legend is overlayed on the x-axis and is tailored to the range of Hedge's *g* for each individual meta-analysis per behavioral test. Dotted line provides a marker for x=0. Bottom diamond represents the random-effects summary effect size and 95% confidence intervals. The underlying cropped color scale represents the prediction interval. Dashed line represents the marker for the summary effect.
Data presentation: bubble plots. Effect size is on the y-axis and levels of the continuous variable moderator are on the x-axis. Color and size scales are the same as the experiment forest plot.

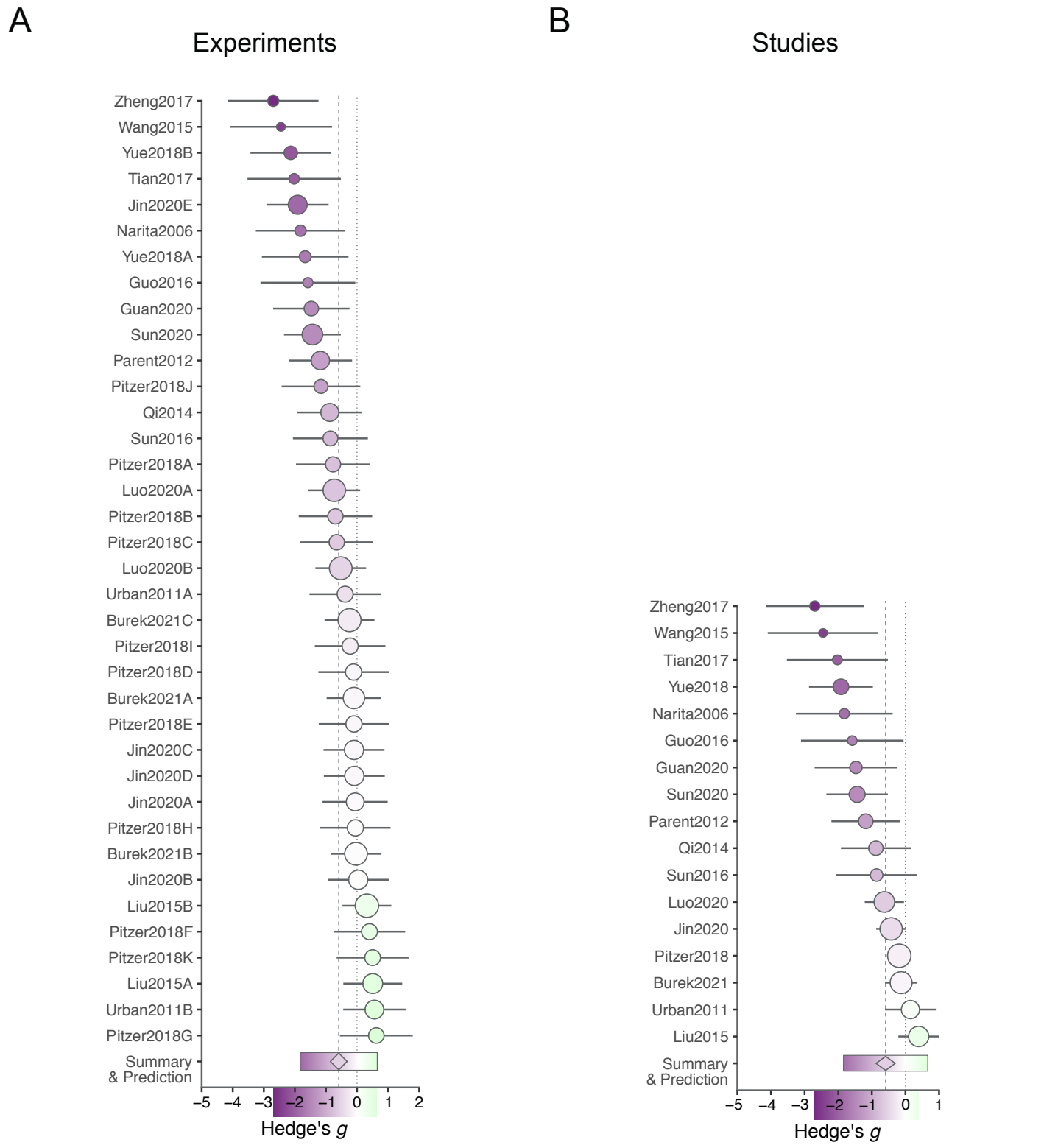


Figure S1. Forest plots of individual and summary estimate standardized mean differences as Hedge's g in elevated plus and zero maze experiments. **A**. Random-effects meta-analysis of experiments comparing open arm time in the EPM/EZM reveals an overall significant CFA-induced reduction in exploratory behavior *g*=-0.5881 [CI –0.8738 to -0.3024; PI -1.8413 to 0.6651], p=0.0002, I^2=54.6%, τ=0.6010, k=37, N=580). **B**. Nesting experiments within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies comparing open arm time in the EPM/EZM maintains an overall significant CFA-induced reduction in exploratory behavior 0.8982 [CI –1.2740 TO -0.5223, PI -2.2755 to 0.4792], p<0.0001, τ=0.6171, I^2=72.3%, k=17, N=580).

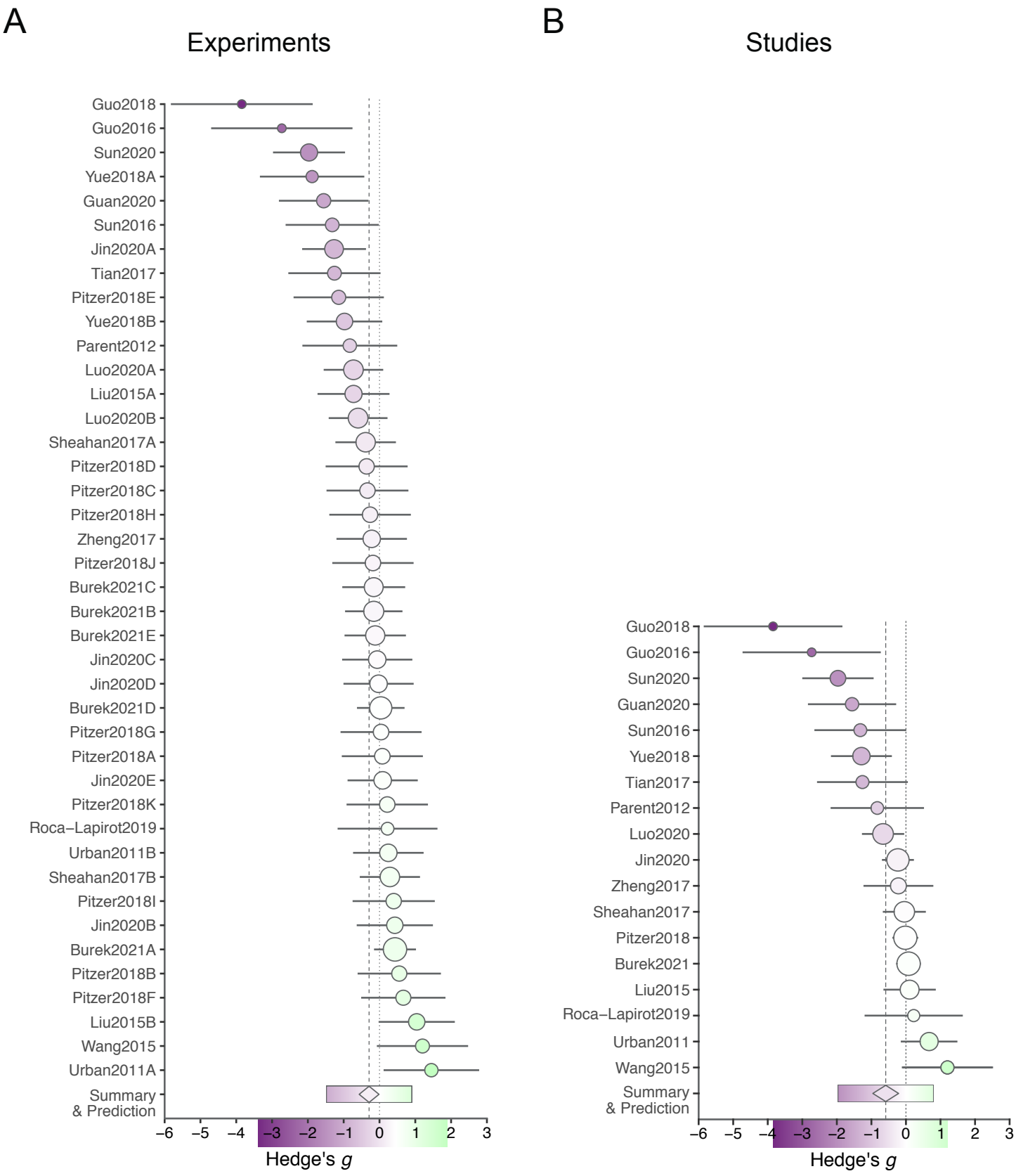


Figure S2. Forest plots of individual and summary estimate standardized mean differences as Hedge's g in open field test experiments. **A**. Random-effects meta-analysis of experiments comparing center time in the OFT reveals an overall significant CFA-induced reduction in exploratory behavior (g=-0.2910 [CI -0.5682 to -0.0138; PI -1.4851 to 0.9031], p=0.0401, I^2=56.2%, τ=0.5742, k=41, N=679). **B**. Nesting experiments within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies comparing center time in the OFT maintains an overall significant CFA-induced reduction in exploratory behavior (g=-0.5237 [CI -0.8837 to -0.1636, PI -1.8623 to 0.8150], p=0.0044, τ=0.6042, I^2=73.8%, k=18, N=679).

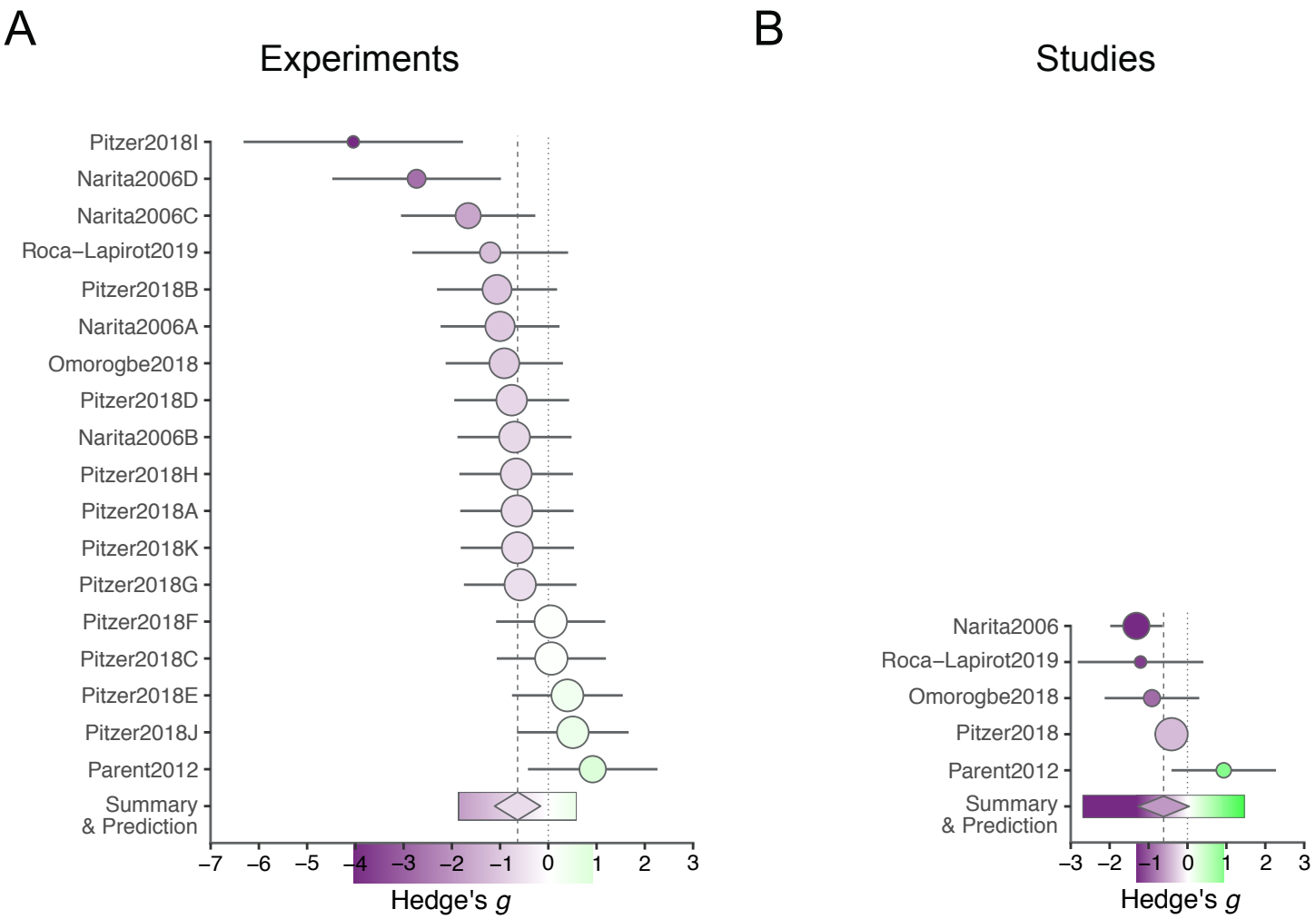


Figure S3. Forest plots of individual and summary estimate standardized mean differences as Hedge's g in light/dark box experiments. **A**. Random-effects meta-analysis of experiments comparing time spent in the light compartment of the LDB reveal an overall significant CFA-induced reduction in exploratory behavior (-0.6369 [-1.1137 to -0.1602], PI [-1.8608 to 0.5870], p=0.0118, τ=0.5313, I^2=47.6%, k=18, N=210). **B**. Nesting experiments within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies comparing time spent in the light compartment of the LDB, indicates a nonsignificant CFA-induced reduction in exploratory behavior (-0.6257 [-1.2826 to 0.0312], PI [-2.7034 to 1.4520], p=0.0619, τ= 0.5603, I^2=63.8%, k=5, N=210).

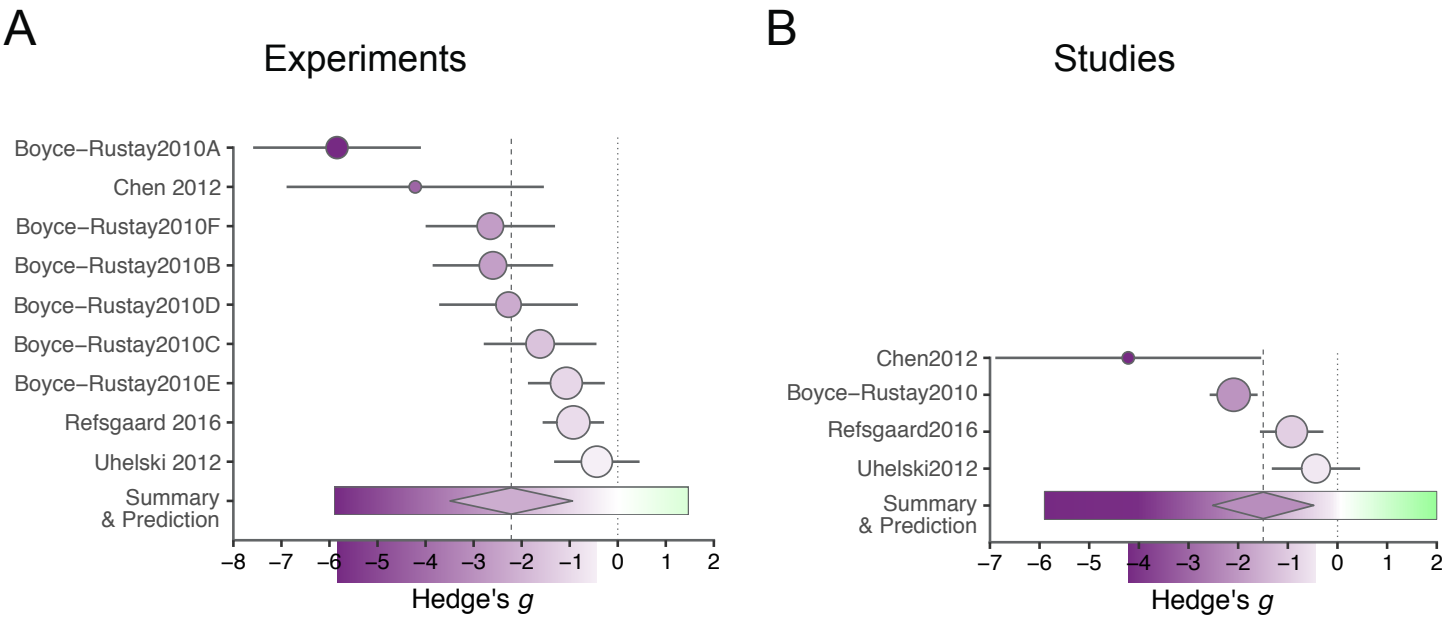


Figure S4. Forest plots of individual and summary estimate standardized mean differences as Hedge's g in place escape/avoidance paradigm experiments. **A**. Random-effects meta-analysis of experiments comparing time spent in a dark compartment paired with noxious hind-paw stimulation reveals an overall significant CFA-induced reduction in time spent (-2.2156 [-3.4905 to -0.9408], PI [-5.9046 to 1.4733], p=0.0039, τ=1.4588, I^2=82.0%, k=9, N=198). **B**. Nesting experiments within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies in the PEAP maintains an overall significant effect of CFA (-1.5048 [-2.5371 to -0.4724], PI [-5.9877 to 2.9782], p=0.0043, τ=0.8990, I^2=83.6%, k=4, N=198).

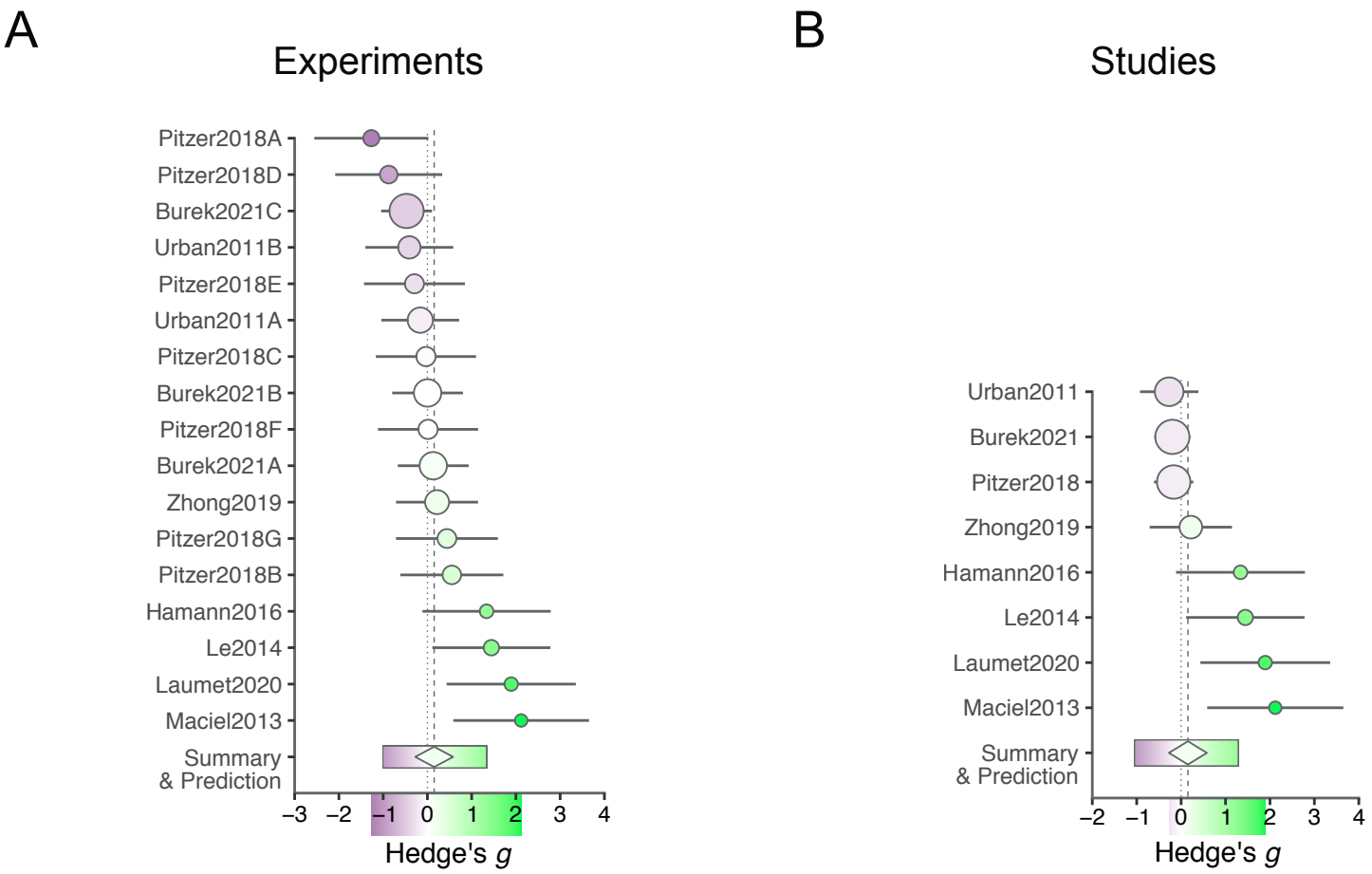


Figure S5. Forest plots of individual and summary estimate standardized mean differences as Hedge's g in forced swim test experiments. **A**. Random-effects meta-analysis of experiments comparing time spent immobile in the FST reveals no overall CFA-induced increase in passive stress coping (*g*=0.4806 [CI -0.0566 to 1.0167; PI -1.1194 to 2.0805], p=0.0789, τ=0.5939, I^2=70.5%, k=8, N=280). **B**. Nesting experiments within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies comparing immobility in the FST also reveals no significant CFA-induced alteration in stress coping strategy (*g*=0.5446 [CI -0.0024 to 1.0916; PI -1.0840 to 2.1732], p=0.0510, τ=0.6042, I^2=67.7%, k=8, N=232).

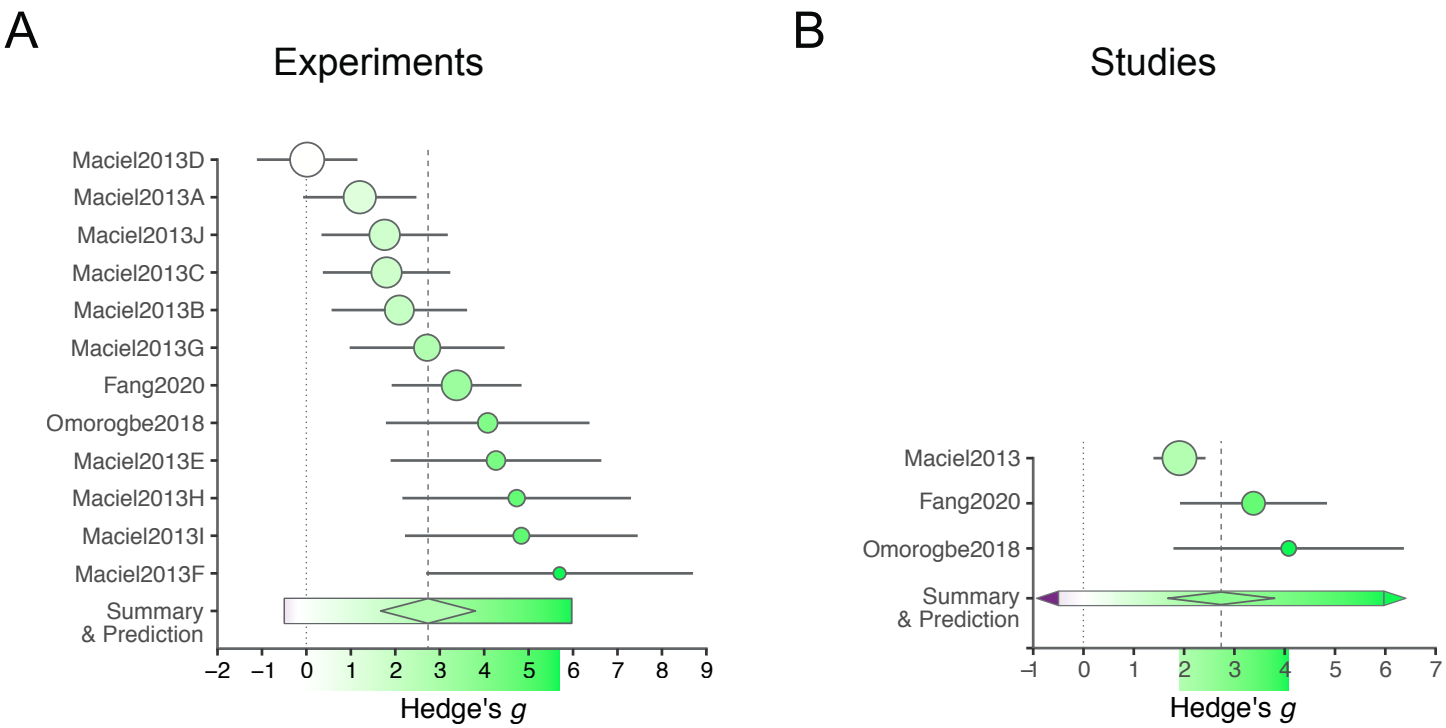


Figure S6. Forest plots of individual and summary estimate standardized mean differences as Hedge's g in tail suspension test experiments. **A**. Random-effects meta-analysis of experiments comparing time spent immobile in the TST reveals an overall significant CFA-induced increase in immobility (2.7387 [CI 1.6723 to 3.8050, PI -0.4949 to 5.9722], p=0.0001, τ=1.3680, I^2=71.4%, k=12, N=152). **B**. Nesting experiments within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies comparing immobility in the TST maintains an overall significant CFA-induced increase (2.8338 [CI 1.4684 to 4.1992, PI -12.4416 to 18.1093], p<0.0001, τ=0.9798, I^2=68.1%, k=3, N=152).

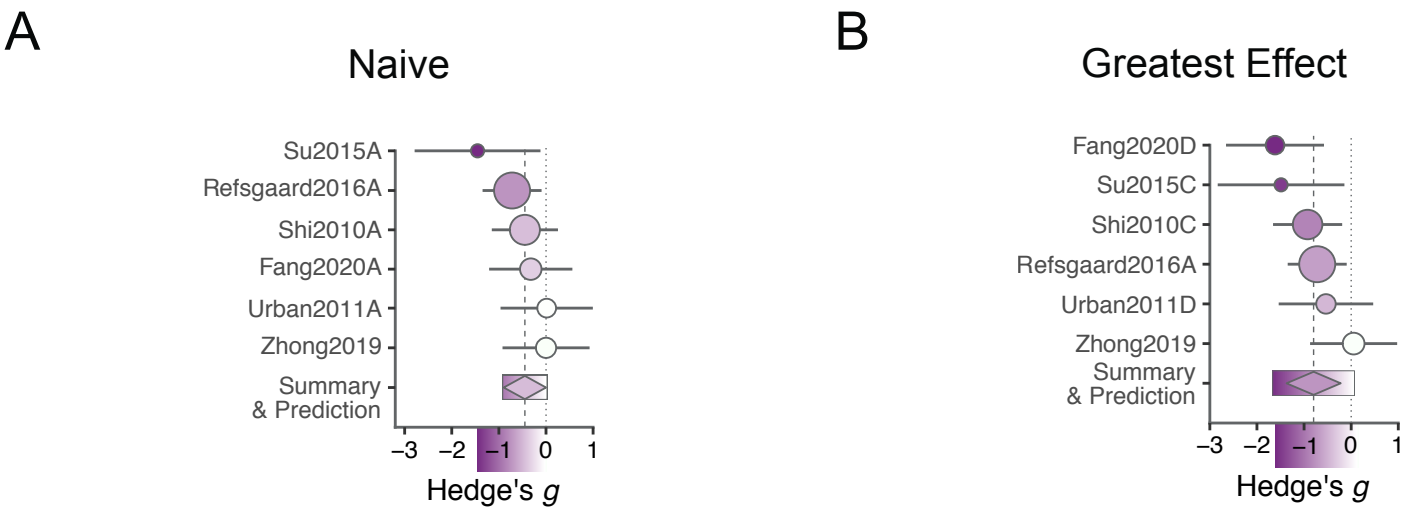


Figure S7. Forest plots of individual and summary estimate standardized mean differences as Hedge's g in sucrose preference experiments. **A**. Random-effects meta-analysis of experiments comparing sucrose preference using only the first of repeated measures, i.e. the first measurement of preference per cohort, reveals an overall significant reduction in sucrose preference by CFA (g=-0.4489 [CI -0.8936 to -0.0041; PI -0.9292 to 0.0315], p=0.0486, I^2<0.0001%, τ=0.0008, k=6, N=140). **B**. Random-effects meta-analysis of experiments comparing sucrose preference using the data point with greatest effect from repeated measures, maintains an overall significant reduction in sucrose preference by CFA (g=-0.7973 [CI -1.3723 to -0.2223; PI -1.6658 to 0.0711], p=0.0161, I^2=29.3%, τ=0.2186, k=6, N=140).

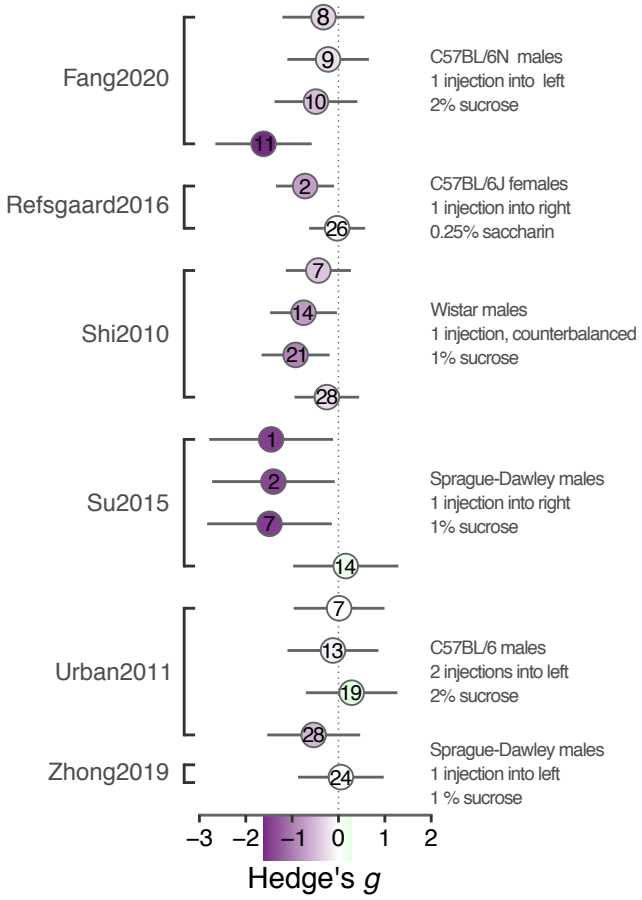


Figure S8. Faceted forest plot of repeated measures per study using sucrose preference. Left brackets delineate unique studies. Text on right provides methodological information on species, strain, sex, number of CFA exposures, side of paw injected, and concentration of sucrose or saccharin used. Number inside circle indicates number of days post-injection when preference measured.

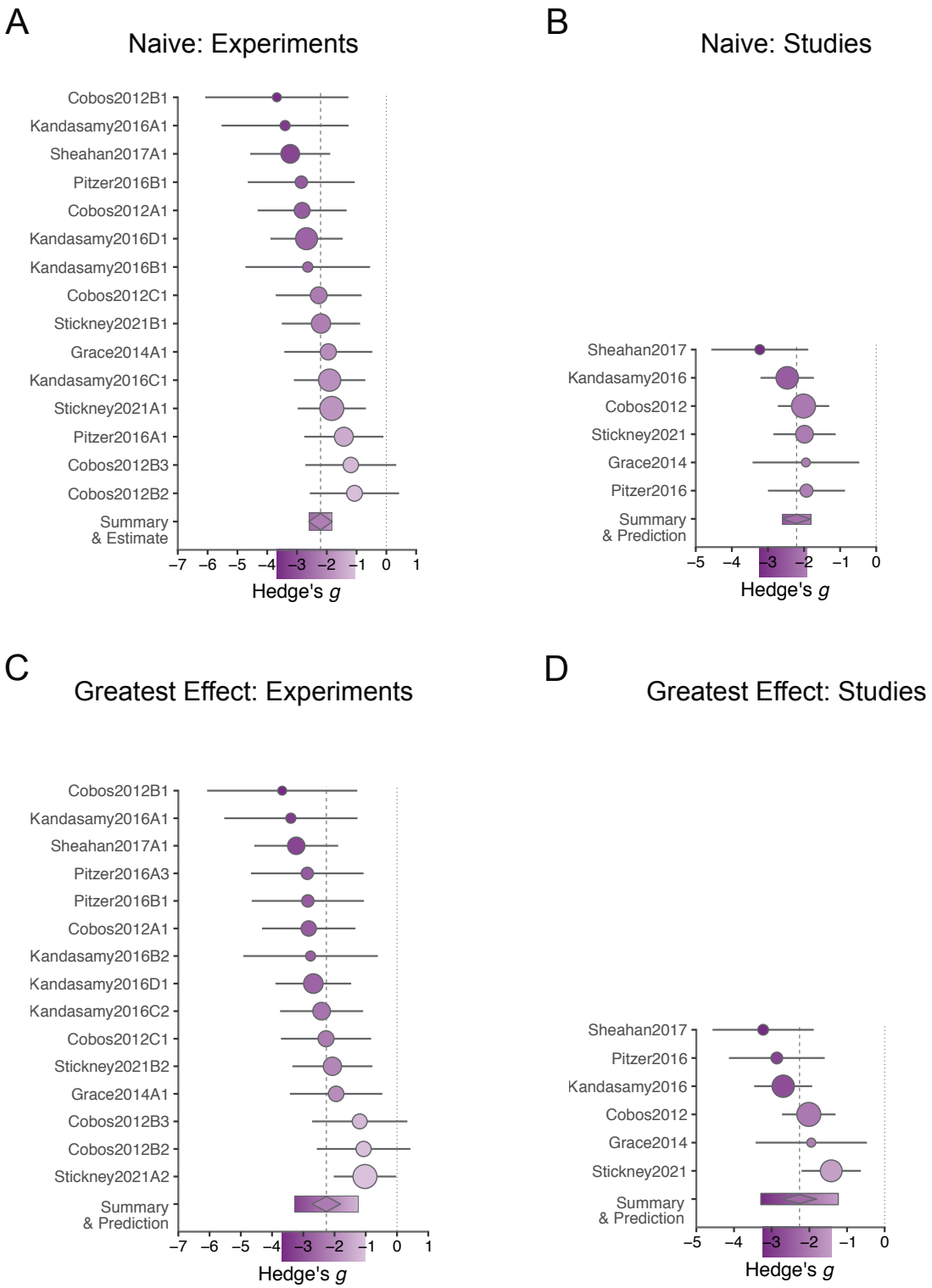


Figure S9. Forest plots of individual and summary estimate standardized mean differences as Hedge's g in wheel running experiments. **A**. Random-effects meta-analysis of experiments comparing wheel running using only the first of repeated measures, i.e. the first measurement of preference per cohort, reveals an overall significant reduction in distance traveled or revolutions by CFA (g=-2.2074 [CI -2.6016 to -1.8132; PI -2.6044 to -1.8103], p<0.0001, I^2<0.00%, τ<0.0001, k=15, N=213). **B**. Nesting "naïve" data within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies maintains an overall significant reduction in distance traveled or revolutions by CFA (g=-2.2074 [CI -2.5845 to -1.8302; PI -2.7416 to -1.6731], p<0.0001, I^2<0.00%, τ<0.0001, k=6, N=213). **C**. Random-effects meta-analysis using the data point with greatest effect from repeated measures, maintains an overall significant effect of CFA (g=-2.2601 [CI -2.7144 to -1.8057; PI -3.2880 to -1.2321], p<0.0001, I^2=17.3%, τ=0.4261, k=15, N=213). **D**. Nesting the "greatest effect" data within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies maintains an overall significant reduction in distance traveled or revolutions by CFA (g=-2.2716 [CI -2.8103 to -1.7330; PI -3.7093 to -0.8339]; p<0.0001, I^2=44.5%, τ=0.4389, k=6, N=213).

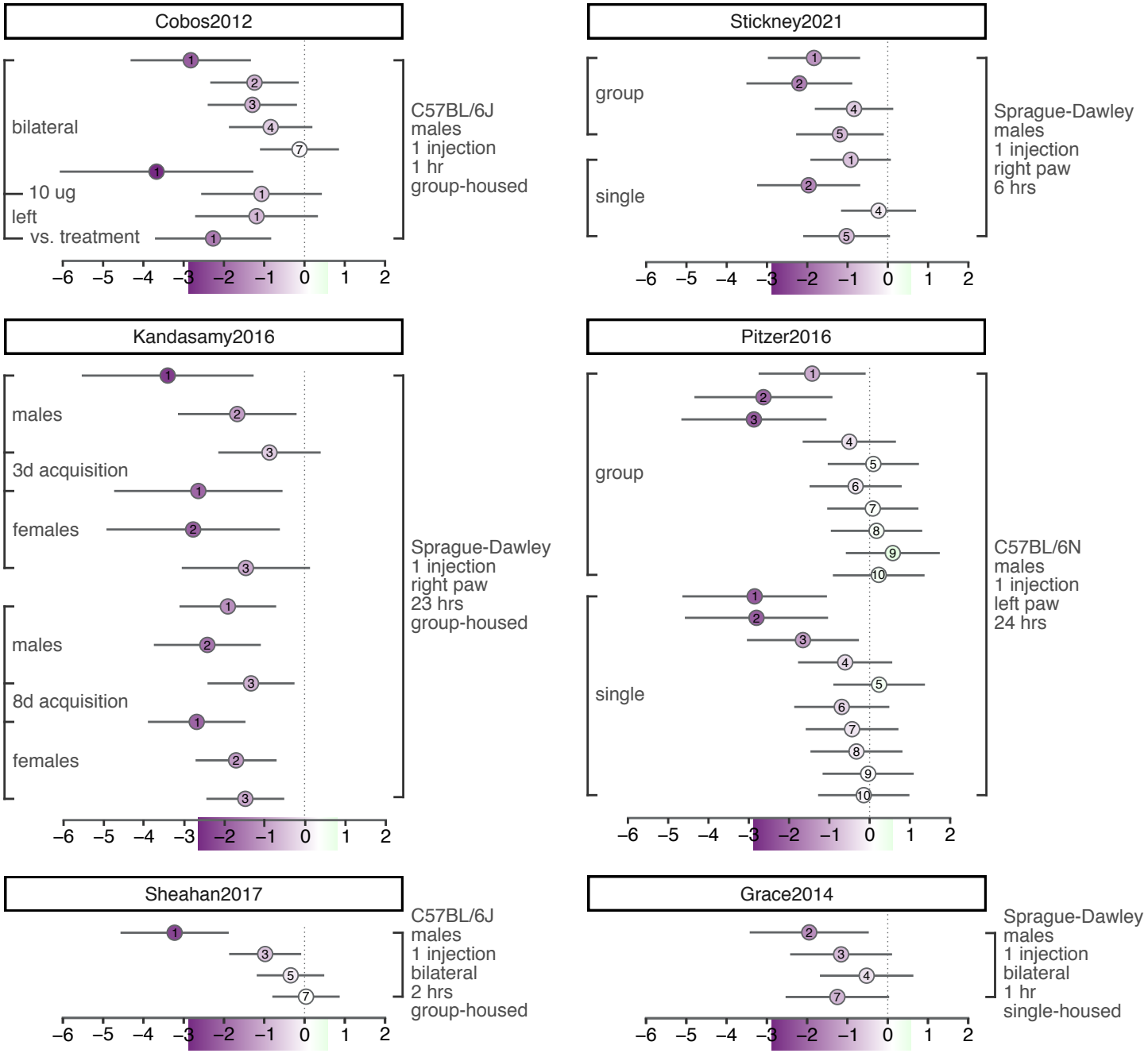


Figure S10. Faceted forest plot of repeated measures per study using wheel running. Top bar per plot indicates unique study. Number inside circle indicates number of days post-injection when preference measured. Text on right provides methodological information on species, strain, sex, number of CFA exposures, side of paw injected, and amount of time during which wheel running was measured. Text on left delineate differences between cohorts.

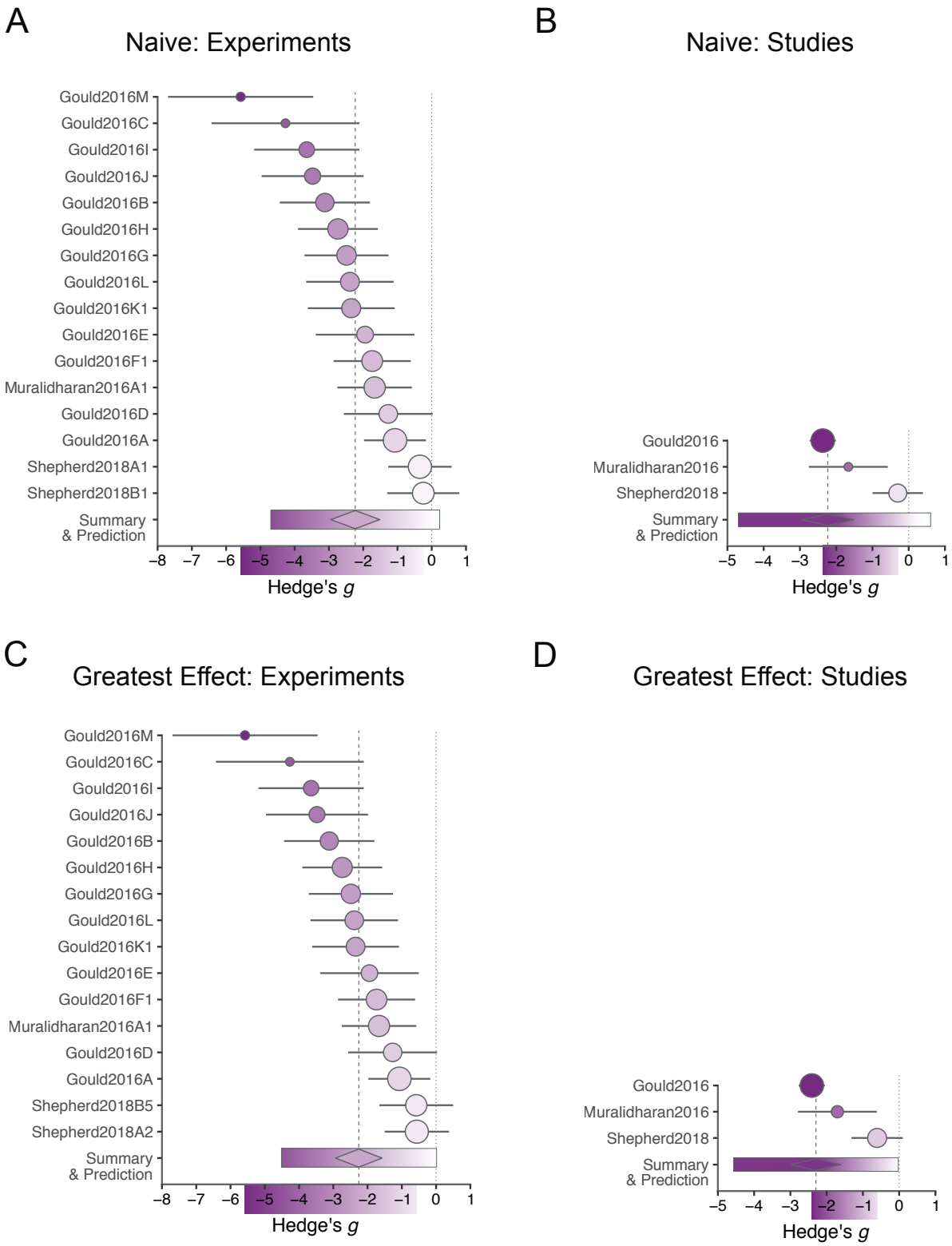


Figure S11. Forest plots of individual and summary estimate standardized mean differences as Hedge's g in burrowing experiments. **A**. Random-effects meta-analysis of experiments comparing burrowing activity using only the first of repeated measures, i.e. the first measurement of preference per cohort, reveals an overall significant reduction in burrowing mass by CFA (g=-2.2323 [CI -2.9460 to -1.5186; PI -4.6967 to 0.232], p<0.0001, I^2=73.9%, τ=1.0992, k=16, N=299.) **B**. Nesting these "naïve" data within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies maintained an overall significant reduction in burrowing by CFA (g=-1.4577 [CI -2.8780 to -0.0375; PI -19.2110 to 16.2956], p=0.0442, τ=1.1946, I^2=92.6%, k-3, N=299). **C**. Random-effects meta-analysis of experiments comparing burrowing activity using the data point with greatest effect from repeated measures maintains an overall significant reduction in burrowing mass by CFA (g=-2.2564, [CI -2.9346 to -1.5781; PI -4.5277 to 0.0149], p<0.0001, I^2=70.26, τ=1.0100, k=16, N=299). **D**. Nesting these "greatest-effects" data within their respective studies in a fixed-effects model, then conducting a random-effects meta-analysis of studies maintained an overall significant reduction in burrowing by CFA (g=-1.5545 [CI -2.7932 to -0.3158; PI -16.8575 to 13.7485], p=0.0139, I^2=90.0%, τ=1.0252, k=3, N=299).

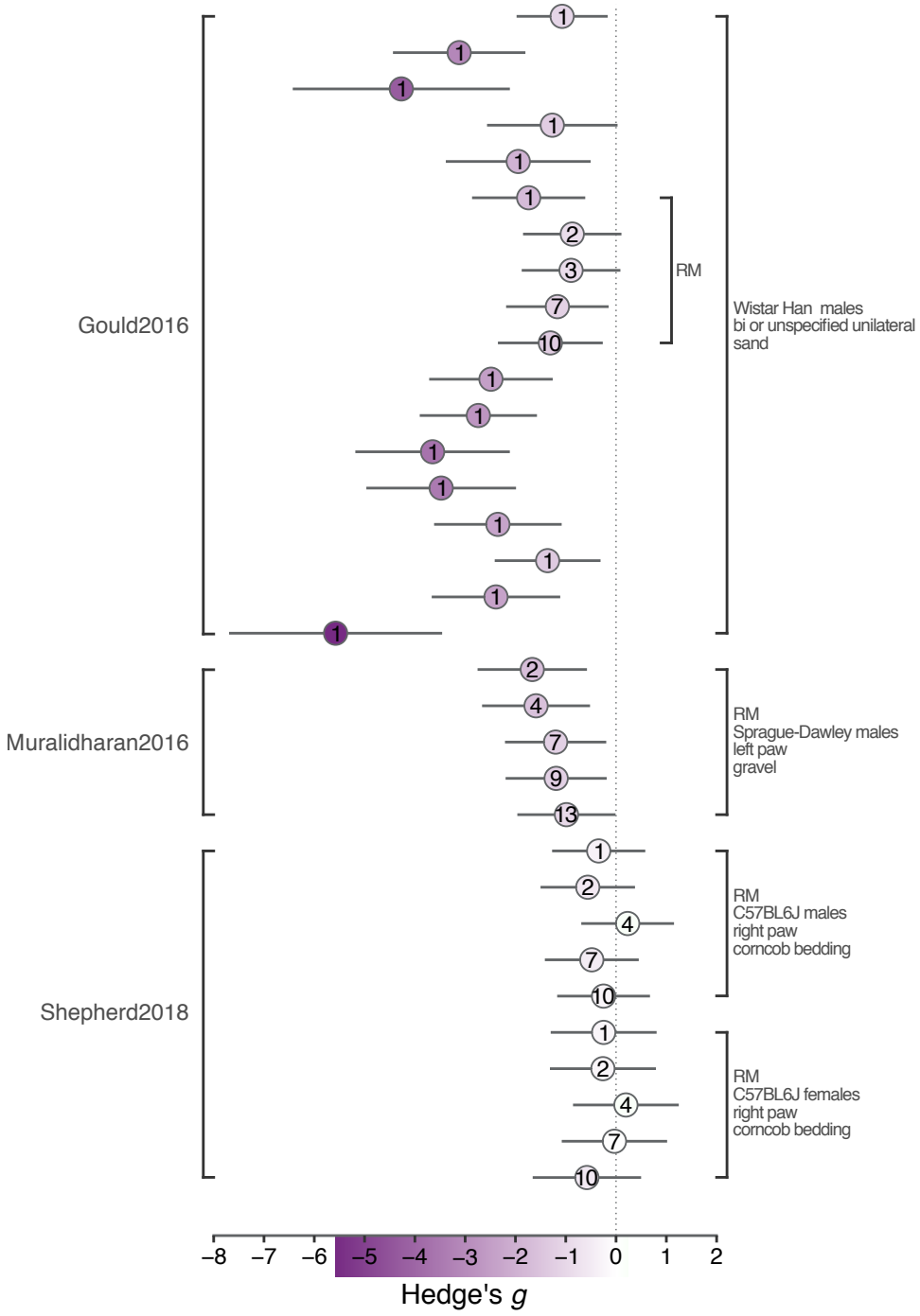
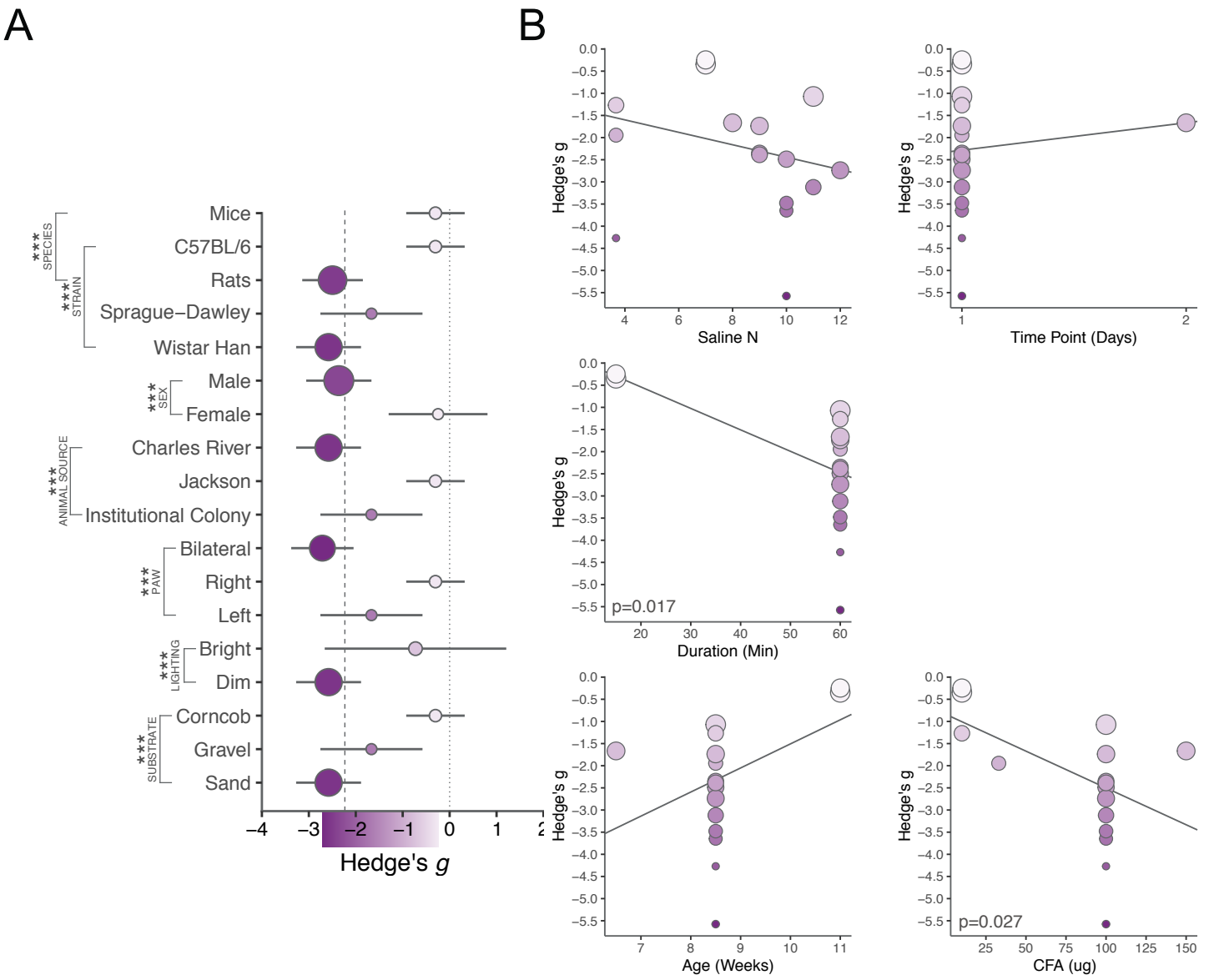
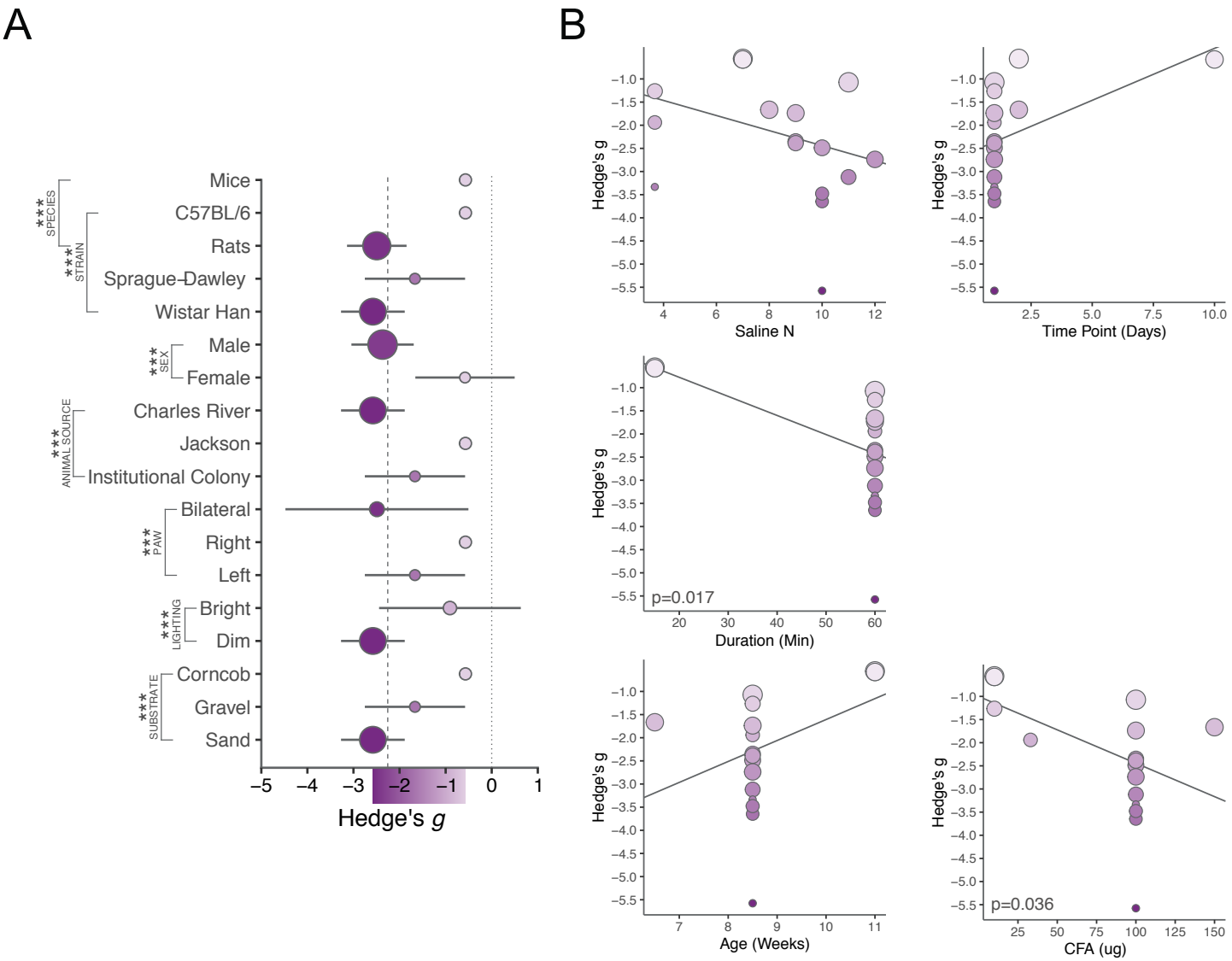


Figure S12. Faceted forest plot of repeated measures per study using burrowing. Left brackets delineate unique studies. Text on right provides methodological information on species, strain, sex, side of paw injected, and substrate used to burrow. Number inside circle indicates number of days post-injection when burrowing was measured. RM bracket indicates cohorts with repeated measures.



S13. Exploring sources of between-experiment heterogeneity in burrowing experiments with subgroup analysis and meta-regression, using the first of repeated measures. **A**. Sub-group analysis of burrowing experiments revealed a significant impact of animal sourcing on CFA-induced deficits, such that animals purchased from Charles River exhibited the greatest burrowing deficit, although only one and two experiments sourced animals from an institutional colony or Jackson Laboratories, respectively (Q_2=55.81, p<0.0001). Strain differences were also significantly apparent, although this was primarily due to group membership overlap in which all Wistar Han rats were purchased from Charles River (Q_2=55.81, p<0.0001). Only one experiment used female rodents, which experienced no significant effect of CFA on their burrowing behavior leading to significant subgroup difference from males (Q_1=11.36, p=0.0008). Bilaterally-injected animals had the greatest burrowing deficit (Q_2=67.55, p<0.0001). Both C57BL/6 mice were of the J substrain. All animals were group-housed, injected just once with CFA purchased from Sigma, and naïve to other testing. Note that only one experiment each used Sprague-Dawley rats, female animals, sourced from an institutional colony, injected into the left paw, or used gravel substrate. **B**. Sample size in the control, saline-injected group was not significantly associated with effect size and accounted for only 5.4% of observed heterogeneity in burrowing deficits ($F_{1,14}$=1.14, p=0.3029). Because the initial trial of any repeated measures were chosen to conduct this meta-analysis, no significant differences were observed between animals tested on day 1 or 2 post-CFA injection but the data are provided for comparison regardless ($F_{1,14}$=0.20, p=0.6581). The amount of CFA, in micrograms, injected into the paw to induce inflammation was a significant moderator of effect, accounting for 39.7% of observed heterogeneity albeit with a significant amount of heterogeneity left over ($F_{1,14}$=6.06, p=0.0274). (N=299).



S14. Exploring sources of between-experiment heterogeneity in burrowing experiments with subgroup analysis and meta-regression, using the greatest effect data point from repeated measures. **A**. All sub-group differences were significant. The four experiments that used rats had a greater reduction in burrowing than the two that used mice (Q_1=41.41, p<0.0001). Wistar Han rats had the greatest reduction in burrowing compared to Sprague-Dawley rats and C57BL/6 mice (Q_2=43.93, p<0.0001). Bilaterally injected animals burrowed less than those injected into either the right or left paw only (Q_2=13.38, p=0.0012). Animals sourced from Charles River, not the institutional colony, burrowed the least (Q_2=43.93, p<0.0001). Although only one study used them, females had a smaller burrowing deficit than males (Q_1=7.97, p=0.0048). Animals burrowed less in dim lighting (Q_1=12.24, p=0.0005). Animals buried sand the least (Q_2=43.93, p<0.0001). Both C57BL/6 mice were of the J substrain. All animals were group-housed and injected just once with CFA purchased from Sigma. Note that only one experiment each used Sprague-Dawley rats, female animals, sourced from an institutional colony, injected into the left paw, or used gravel substrate. **B**. Duration of access to substrate for burrowing was a significant moderator of effect, accounting for 53.3% of heterogeneity ($F_{1,14}$=7.33, p=0.0170). Amount of CFA injected was also a significant moderator of effect, accounting for 37.8% of heterogeneity ($F_{1,14}$=5.39, p=0.0358). Age ($F_{1,14}$=2.96, p=0.1073), days after CFA injection ($F_{1,14}$=3.43, p=0.0852), and sample size in the control group ($F_{1,14}$=1.95, p=0.1848) were not significant moderators.

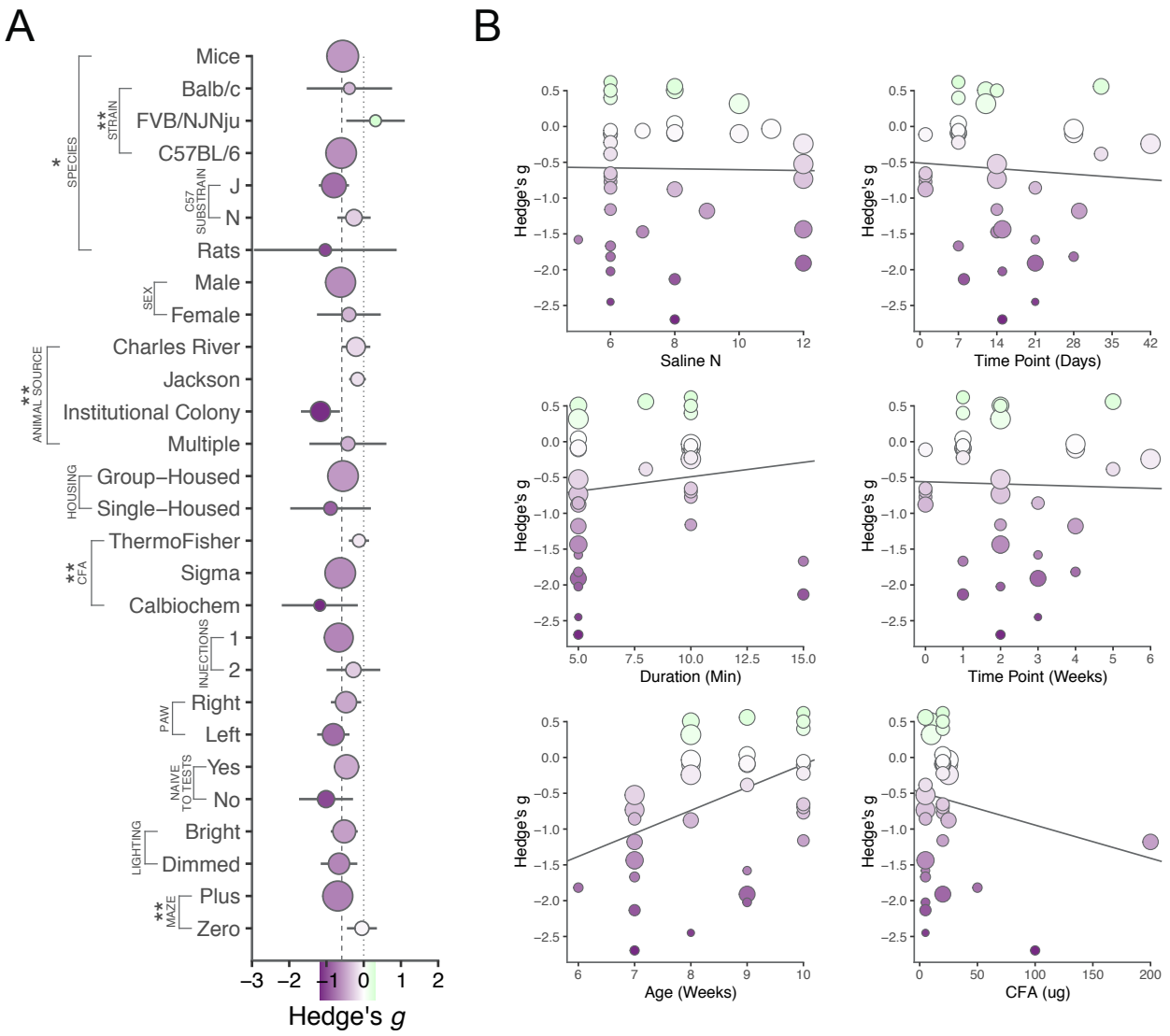


Figure S15. Exploring sources of between-experiment heterogeneity in elevated plus or zero maze experiments with subgroup analysis and meta-regression. **A**. CFA significantly reduces open arm time to a greater extent in rats than in mice (Q_1=4.87, p=0.0273). Open arm time also differs by strain, although most experiments utilized C57BL/6 mice, only one experiment each used Balb/c or FVB/NJNju mice, and all three rat studies used Sprague-Dawley rats (Q_3=11.80, p=0.0081). Among experiments where such information was available, C57 substrain differences were very nearly significant (Q_1=3.82, p=0.0506). Animals sourced from institutional colonies rather than purchased from a vendor such as Charles River or Jackson exhibited the greatest reductions in open arm time after CFA injection (Q_3=15.61, p=0.0014). Most experiments sourced CFA from Sigma-Aldrich. Experiments using ThermoFisher CFA were overall nonsignificant, and one experiment using Calbiochem CFA observed the greatest reduction in exploratory behavior (Q_2=12.30, p=0.0021. Experiments using a plus instead of a zero maze observed significant reduction in exploratory behavior after CFA-injection (Q_1=9.16, p=0.0025). Note that only one experiment each used Balb/c mice, FVB/NJNju mice, or CFA from Calbiochem. **B**. Sample size in the control group accounts for 0.0% of observed heterogeneity ($F_{1,35}$=0.01, p=0.9231). Interval between CFA-injection and assessment in neither days nor weeks accounts for any observed heterogeneity ($F_{1,36}$= 0.30, p=0.5879; $F_{1,35}$=0.03, p=0.8697). Duration of assessment in the maze accounted for 0.0% of heterogeneity as well ($F_{1,35}$= 0.66, p=0.4233). Amount of CFA injection in micrograms accounted for 1.8% of heterogeneity but was not a significant moderator of effect ($F_{1,35}$=1.26, p=0.2694). Age at CFA injection, in weeks, was a significant moderator of effect, accounting for 30.4% of observed heterogeneity ($F_{1,35}$=9.16, p=0.0046).

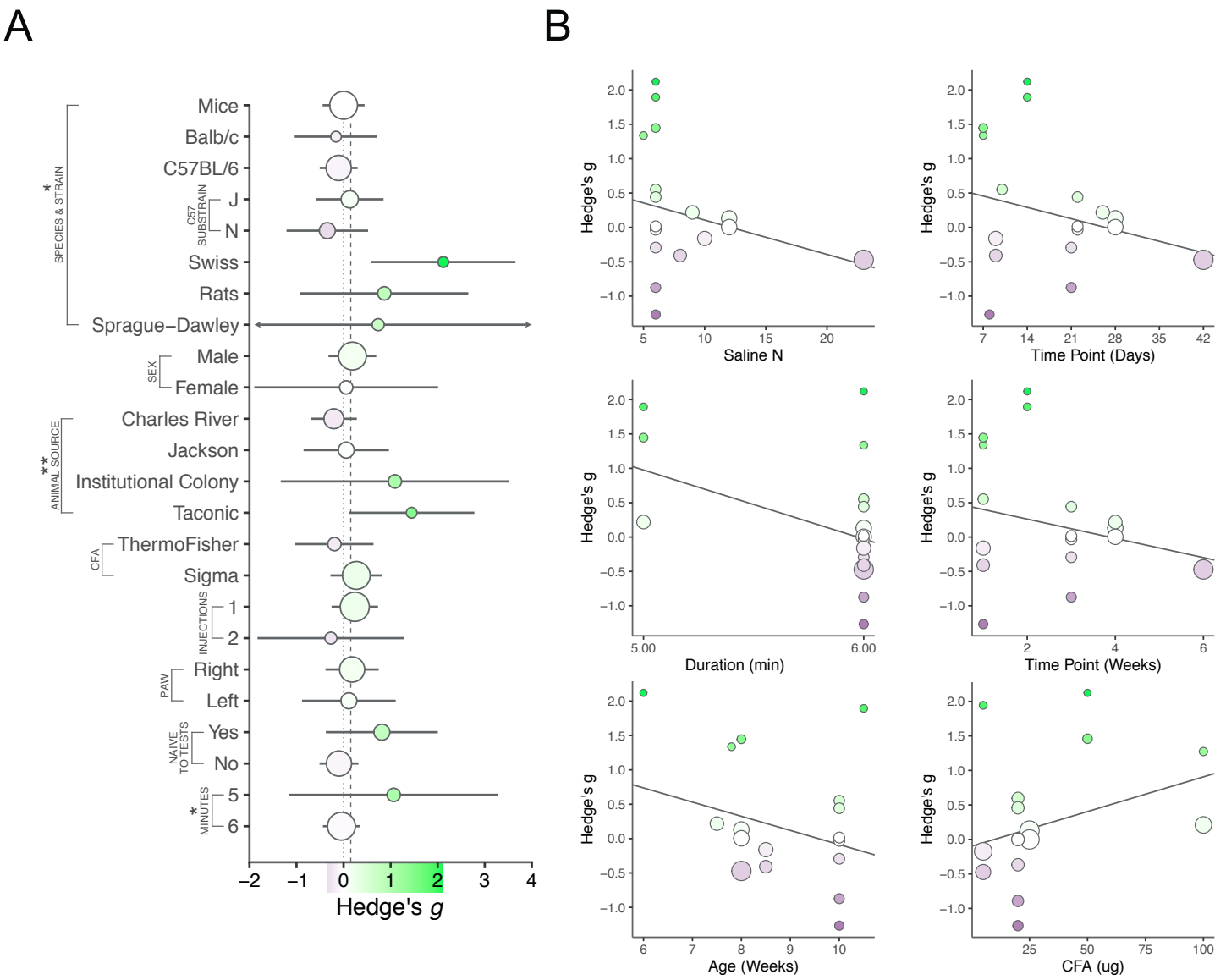


Figure S16. Exploring sources of between-experiment heterogeneity in forced swim test experiments with subgroup analysis and meta-regression. **A**. Immobility in the FST was also significantly affected by animal sourcing, such that the greatest increase in immobility was observed in the one experiment using rodents sourced from Taconic, with no significant change in immobility in the majority of experiments using Charles River (Q_3=9.17, p=0.0272). Strain differences were also observed in CFA-induced forced swim immobility; the majority of experiments used C57BL/6 mice which overall had the lowest magnitude of effect (Q_4=12.14, p=0.0164). Sub-group differences were also apparent between experiments analyzing 5 vs. 6 minutes (Q_1=4.11, p=0.0426). **B**. Amount of CFA injected accounted for 18.9% of heterogeneity but was not a significant predictor of treatment effect ($F_{1,15}$=2.16, p=0.1620). Sample size in the control group accounted for 19.1% of heterogeneity, but was not significant ($F_{1,15}$=1.74, p=0.2067). Interval to assessment, in days, accounted for 9.2% of heterogeneity but was not a significant moderator of effect ($F_{1,15}$=1.52, p=0.2365). Neither was interval to assessment in weeks (3.1%, $F_{1,15}$=1.19, p=0.2920). Age accounted for 0.0% heterogeneity ($F_{1,15}$=1.41, p=0.2540).

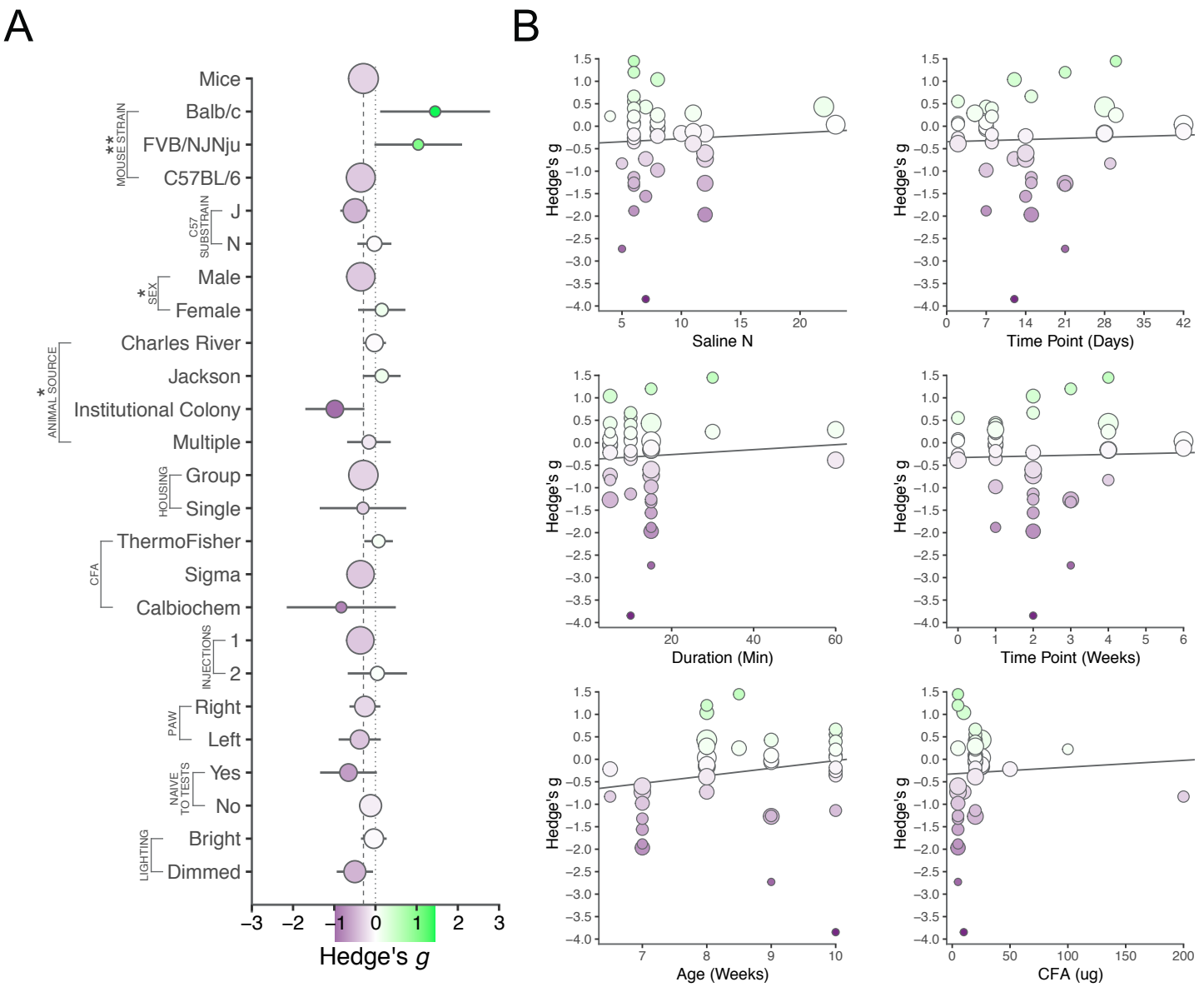


Figure S17. Exploring sources of between-experiment heterogeneity in open field test experiments with subgroup analysis and meta-regression. **A**. Center time in the OFT after CFA differs by strain, although most experiments utilized C57BL/6 mice, only one experiment each used Balb/c or FVB/NJNju mice, and two rat studies used Sprague-Dawley rats (Q_3=12.40, p=0.0061). Animals sourced from institutional colonies rather than purchased from a vendor such as Charles River or Jackson exhibited the greatest reductions in open arm time after CFA injection (Q_3=9.46, p=0.0238). Although most experiments used males with only five using females, females overall had a very slight nonsignificant increase in center time compared to a slight significant decrease in males (Q_1=3.91, p=0.0480). Note that only one experiment each used FVB/NJNju mice, Balb/c mice, or CFA from Calbiochem. **B**. Sample size in the control group accounts for 0.0% of observed heterogeneity ($F_{1,39}$= 0.18, p=0.6766). Interval between CFA-injection and assessment in neither days nor weeks accounts for any observed heterogeneity ($F_{1,39}$=0.08, p=0.7837; $F_{1,39}$=0.04, p=0.8397). Duration of assessment in the maze accounted for 0.0% of heterogeneity as well ($F_{1,39}$= 0.25, p=0.6172). Amount of CFA injection in micrograms accounted for 7.4% of heterogeneity but was not a significant moderator of effect ($F_{1,39}$= 0.10, p=0.7539). Age at CFA injection, in weeks, was not significant moderator of effect, accounting for 11.2% of observed heterogeneity ($F_{1,39}$= 1.99, p=0.1663).

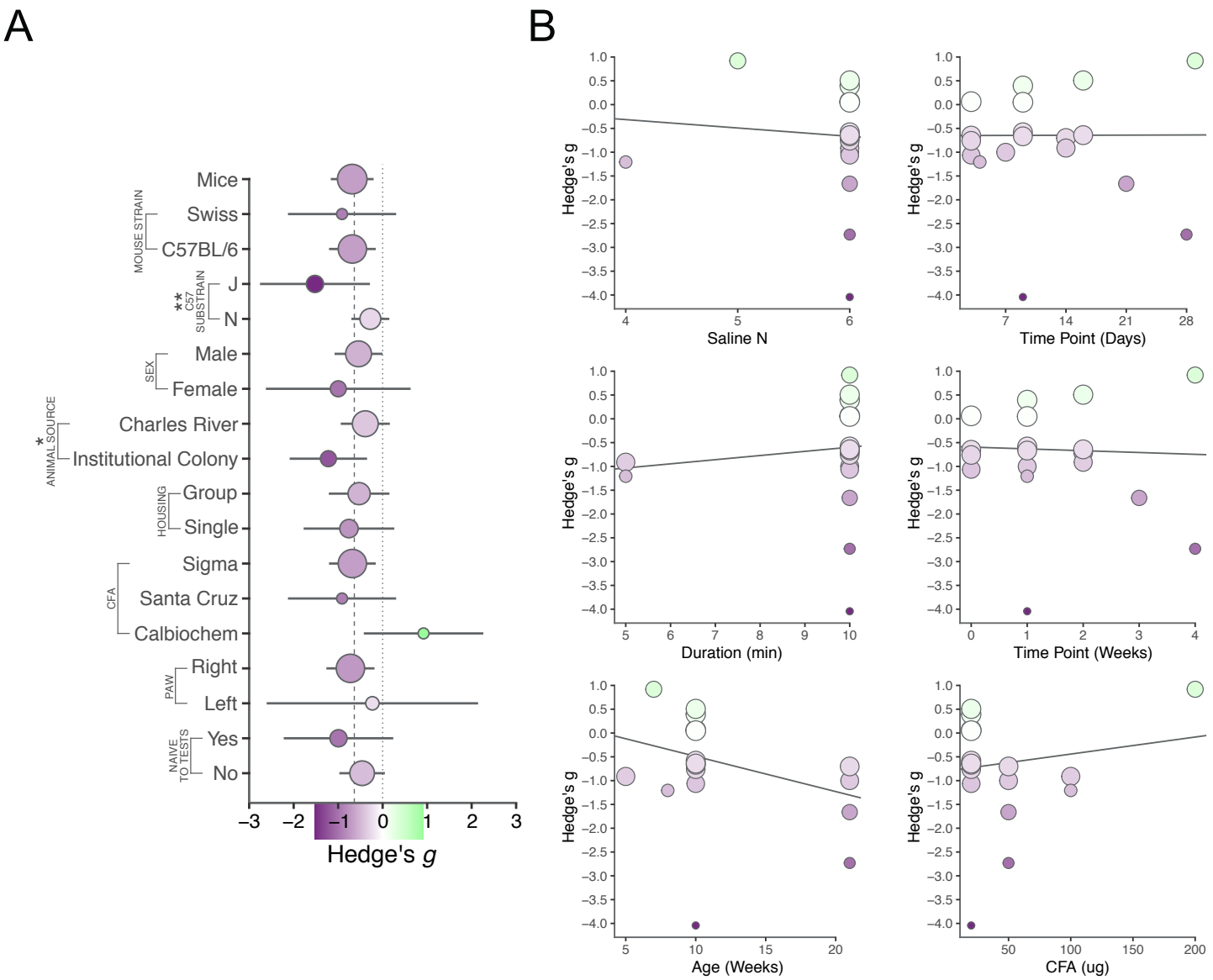


Figure S18. Exploring sources of between-experiment heterogeneity in light/dark box experiments with subgroup analysis and meta-regression. **A**. Animals sourced from institutional colonies rather than purchased from Charles River exhibited the greatest reductions in light compartment time after CFA injection (Q_1=4.24, p=0.0395). C57BL/6 mice of the J sub-strain also exhibited a greater effect of CFA on time spent in the light compartment than those of the N sub-strain (Q_1=5.85, p=0.0156). All animals were injected with CFA once. Note that only one experiment each used Swiss mice, CFA from Santa Cruz, or CFA from Calbiochem. **B**. Sample size in the control group accounts for 0.0% of observed heterogeneity ($F_{1,16}$=0.11, p=0.7402). Interval between CFA-injection and assessment in neither days nor weeks accounts for any observed heterogeneity ($F_{1,16}$=0.0002, p=0.9887; $F_{1,16}$=0.04, p=0.8524). Duration of assessment in the maze accounted for 0.0% of heterogeneity as well ($F_{1,16}$=0.32, p=0.5788). Amount of CFA injection in micrograms accounted for 0.0% of heterogeneity ($F_{1,16}$=0.47, p=0.5020). Age at CFA injection, in weeks, was a significant moderator of effect, accounting for 43.6% of observed heterogeneity ($F_{1,16}$=3.00, p=0.1027).

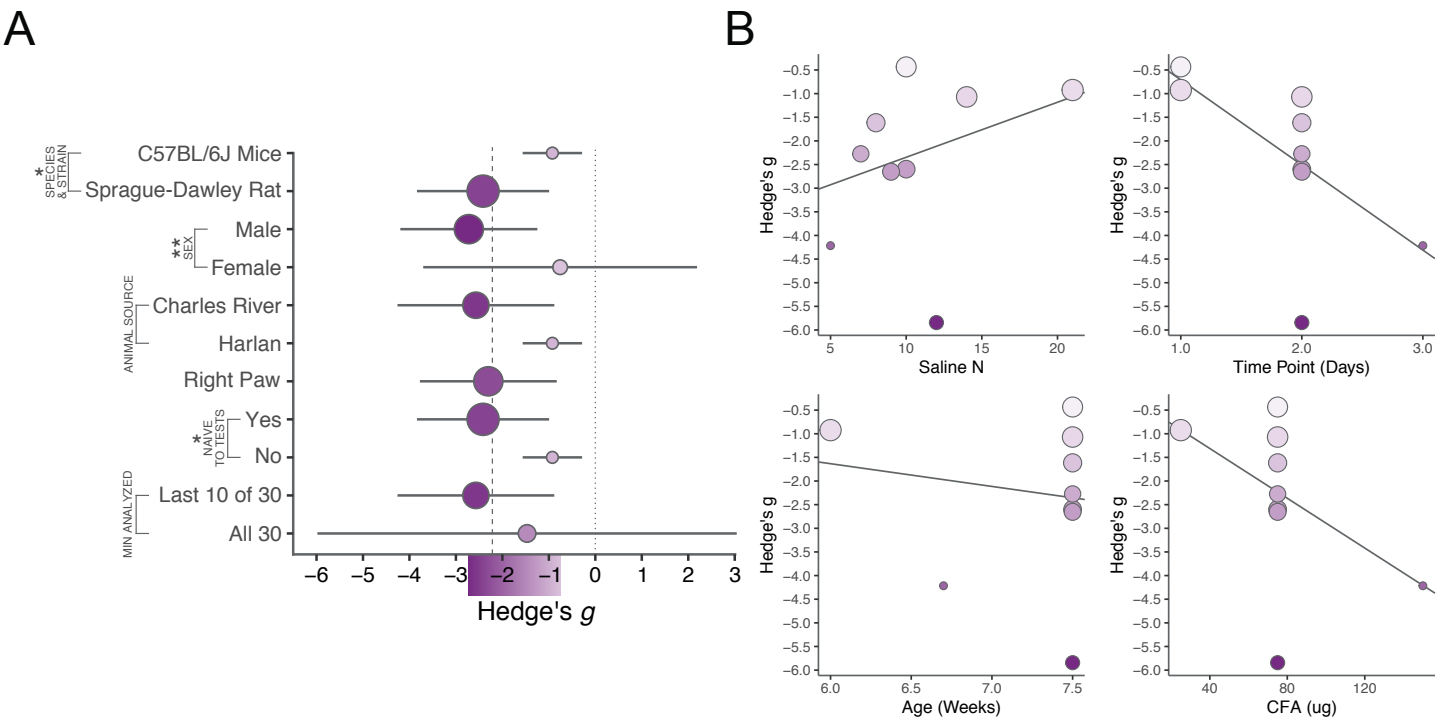


Figure S19. Exploring sources of between-experiment heterogeneity in place escape/avoidance paradigm experiments with subgroup analysis and meta-regression. **A**. Experiments were conducted mostly using either Sprague-Dawley rats; 1 experiment used C57BL/6J mice, making strain and species subgroups significantly different (Q_1=4.75, p=0.0292). Sex differences were also apparent, although this was related to most studies in males being conducted with Sprague-Dawley rats (Q_1=9.25, p=0.0024). All animals were group-housed, injected once with Sigma CFA, and only one C57 mouse was used. Note that only one experiment used mice, animals purchased from Harlan, or had animals assessed previously in other behavioral assays. **B**. Micrograms of CFA injected explain 11.8% of heterogeneity but did not attain significance as a moderator ($F_{1,7}$=1.81, p=0.2209). Interval between CFA injection and PEAP assessment was very nearly a significant moderator of effect, explaining 41.7% of heterogeneity ($F_{1,7}$=4.56, p=0.0700 TPD). Neither age ($F_{1,7}$=0.18, p=0.6817) nor sample size in the control group ($F_{1,7}$=0.88, p=0.3797) explained any heterogeneity. Bin (last 10 min vs. total 30 min) was analyzed in sub-group analysis instead of as Duration. All experiments were conducted less than one week post-CFA injection.

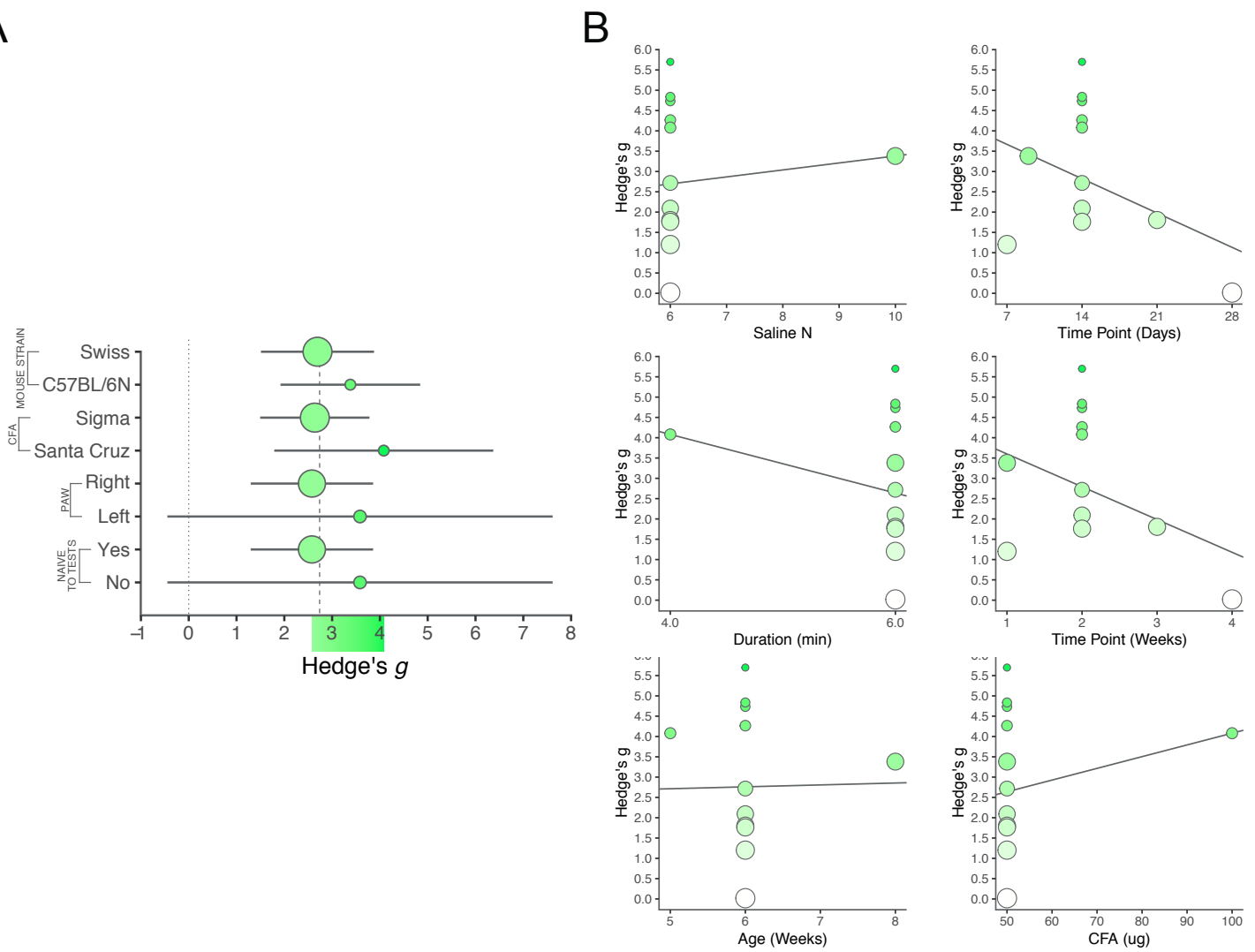


Figure S20. Exploring sources of between-experiment heterogeneity in tail suspension test experiments with subgroup analysis and meta-regression. **A**. No subgroup differences were significant, mostly due to 10 out of 12 experiments coming from the same study with similar methodology. All experiments were conducted using male mice, group-housed, sourced from the institutional colony, and injected just once with CFA. Only one experiment used C57BL/6 mice, of the N substrain. Note that only one experiment each used Santa Cruz CFA or a 4-minute testing duration. **B**. Interval between CFA injection and assessment in the TST is a significant moderator of effect in both days and weeks, accounting for about 17.0% of observed heterogeneity ($F_{1,10}$=2.24, p=0.1657; $F_{1,10}$=2.27, p=0.1629). Age ($F_{1,10}$=0.004, p=0.9527), micrograms of CFA injected and duration of testing ($F_{1,10}$=0.57, p=0.4687), and sample size in the saline group ($F_{1,10}$=0.16, p=0.6988) accounted for no heterogeneity.

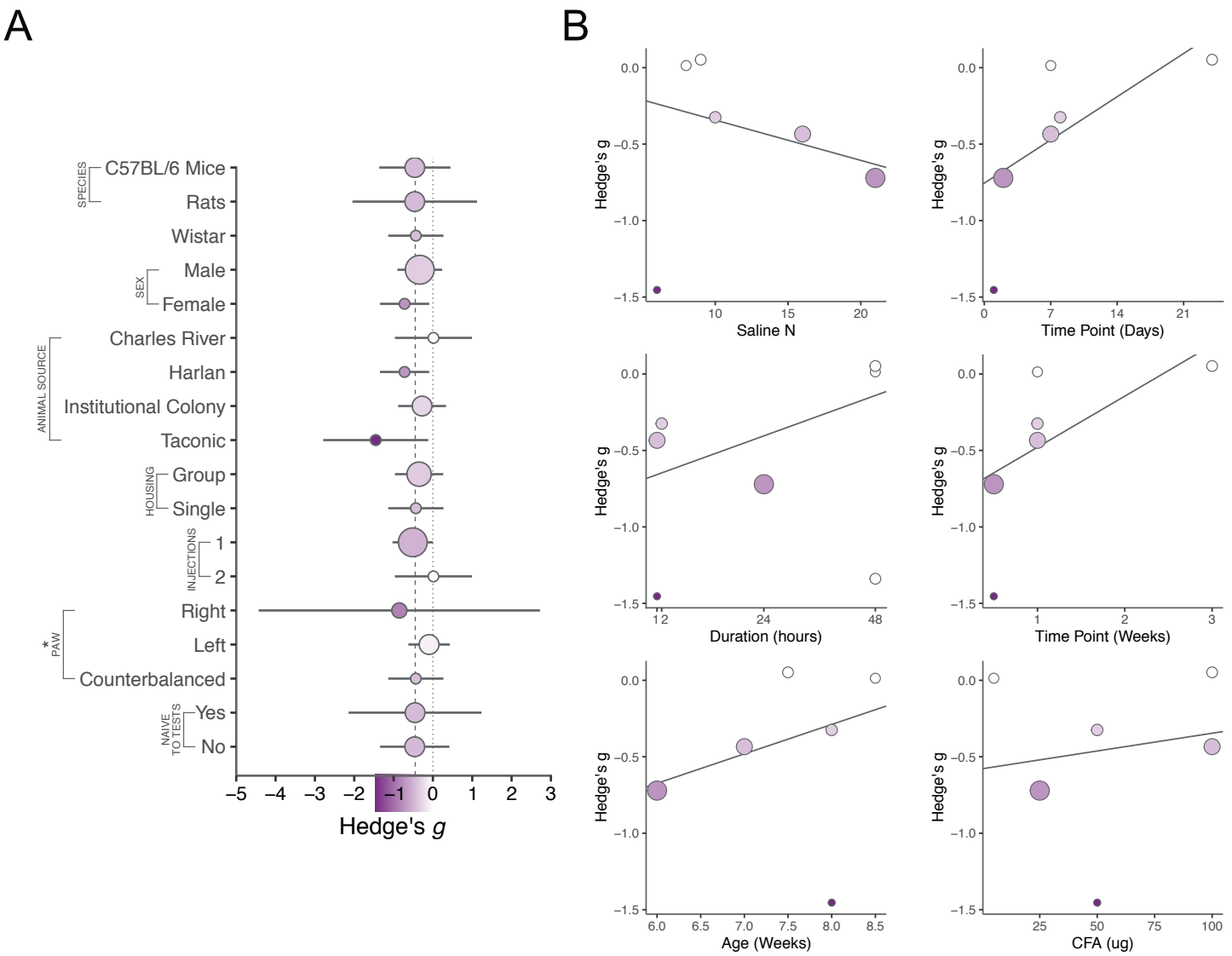


Figure S21. Exploring sources of between-experiment heterogeneity in sucrose preference experiments with subgroup analysis and meta-regression, using the first of repeated measures. **A**. Animals injected into the right paw exhibited more reduced sucrose preference than those injected into the left paw (Q_2=6.39, p=0.0410). All animals were injected with CFA from Sigma. Note that only one experiment each used Wistar rats, C57BL/6J mice, C57BL/6N mice, female animals, animals sourced from Harlan, animals sourced from Taconic, animals sourced from Charles River, single-housing, two CFA injections, or a counter-balanced design for which paw was injected. **B**. No variable was found to be a significant moderator of effect (Age $F_{1,4}$=0.97, p=0.3813; ug CFA $F_{1,4}$=0.20, p=0.6778; Duration $F_{1,4}$=1.66, p=0.2673; Time Point Weeks $F_{1,4}$=3.49, p=0.1349; Time Point Days $F_{1,4}$=4.52, p=0.1006; Saline N $F_{1,4}$=0.64, p=0.4687).

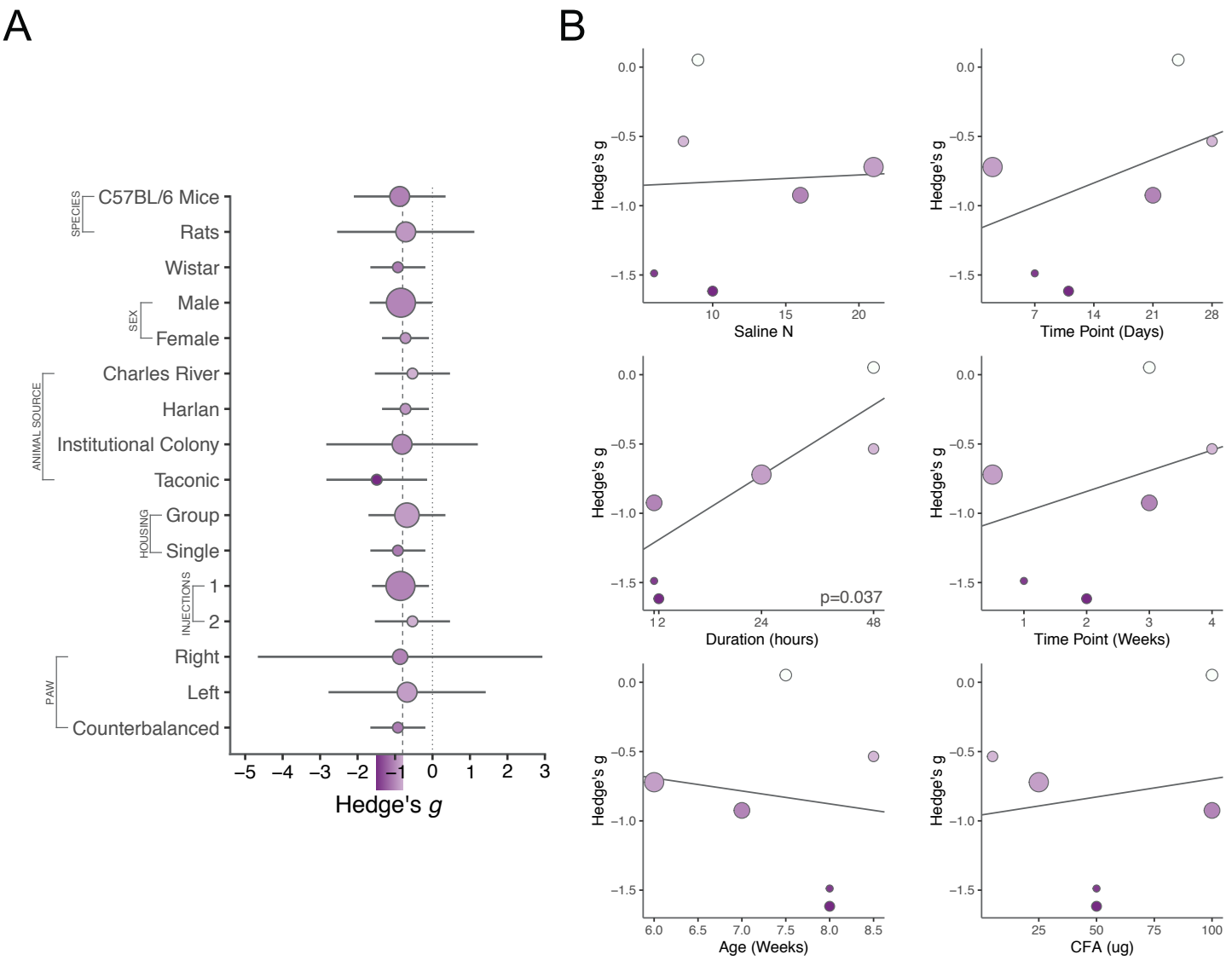


Figure S22. Exploring sources of between-experiment heterogeneity in sucrose preference experiments with subgroup analysis and meta-regression, using the data point with greatest effect from repeated measures. **A**. No significant subgroup differences were identified. All animals were injected with CFA from Sigma. Note that only one experiment each used Wistar rats, C57BL/6J mice, C57BL/6N mice, female animals, animals sourced from Harlan, animals sourced from Taconic, animals sourced from Charles River, single-housing, two CFA injections, or a counter-balanced design for which paw was injected. **B**. Duration of access to sucrose prior to preference measurement was a significant moderator of effect ($F_{1,4}$=9.43, p=0.0372). No other variable was found to be a significant moderator of effect (Age $F_{1,4}$=0.01, p=0.770; ug CFA $F_{1,4}$=0.14, p=0.7299; Time Point Weeks $F_{1,4}$=0.58, p=0.4889; Time Point Days $F_{1,4}$=1.05, p=0.3635; Saline N $F_{1,4}$=0.01, p=0.9219).

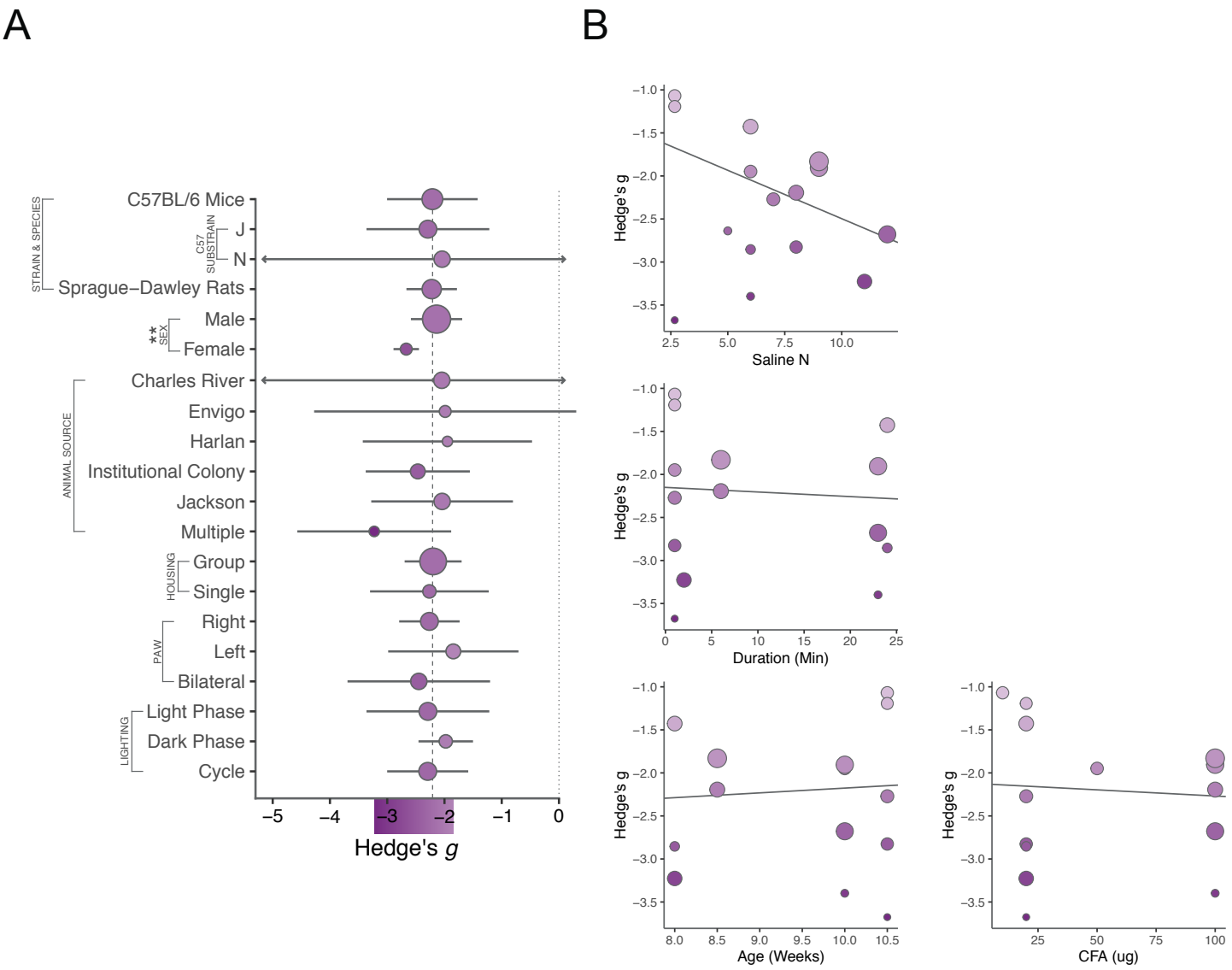


Figure S23. Exploring sources of between-experiment heterogeneity in wheel running experiments with subgroup analysis and meta-regression, using only the first of repeated measures. **A**. Significant differences were observed by sex such that females exhibited a greater CFA-induced deficit in burrowing (Q_1=6.59, p=0.01028). All animals were injected just once with CFA from Sigma. Note that only one experiment each used animals from multiple sources or from Harlan. **B**. Age of animal accounted for no heterogeneity ($F_{1,13}$=0.09, p=0.7735). Micrograms of CFA injected was not a significant moderator of effect ($F_{1,13}$=0.09, p=0.771). Duration of time for which animals were given access to a running wheel prior to measurement was also not a significant moderator of effect ($F_{1,13}$=0.08, p=0.7786). Sample size in the control group accounted for no heterogeneity ($F_{1,13}$=3.32, p=0.0914). Most animals were tested 1 day post-CFA injection, except for 1 cohort tested 2 days later.

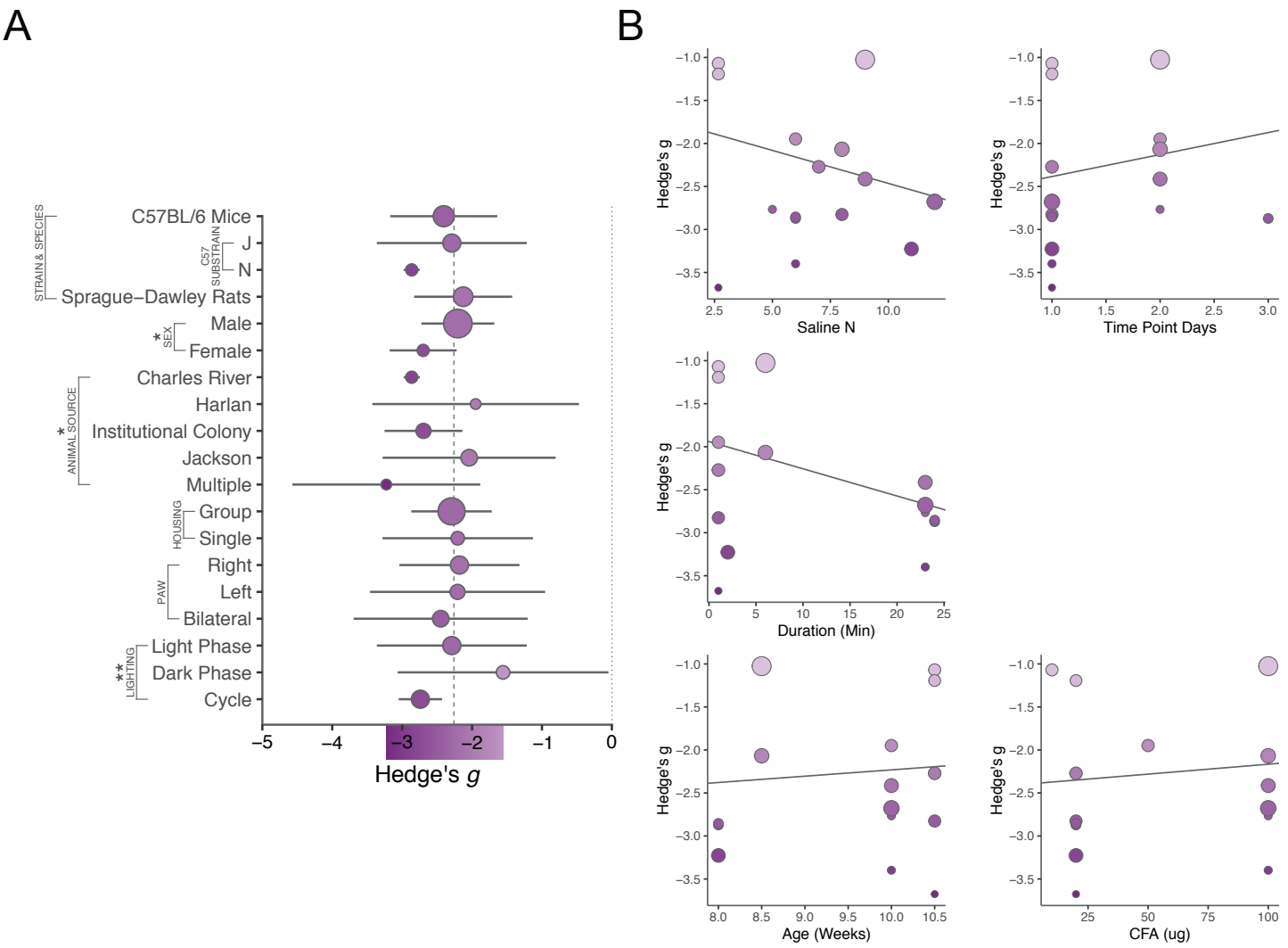


Figure S24. Exploring sources of between-experiment heterogeneity in wheel running experiments with subgroup analysis and meta-regression, using data with the greatest effect from repeated measures. **A**. Significant differences were observed by sex such that females exhibited a greater CFA-induced deficit in burrowing (Q_1=4.19, p=0.0406). Subgroup differences were also observed by animal source, with animals sourced from Harlan and Jackson running the most and animals from institutional colonies and multiple sources running the least (Q_4=13.42, p=0.0198). Animals allowed to run on the wheel across an entire light cycle exhibited the greatest effect of CFA on reduced wheel running, whereas animals tested in the dark phase exhibited the lowest effect (Q_2=10.76, p=0.0041). All animals were injected once with Sigma CFA. Note that only one experiment each used animals from multiple sources or from Harlan. **B**. Duration of access to a running wheel accounted for 25.1% of heterogeneity but was not a significant moderator of effect ($F_{1,13}$=2.43, p=0.1429). Days after CFA injection at which running was measured accounted for only 1.2% of heterogeneity ($F_{1,13}$=0.49, p=0.4959). Sample size in the control group accounted for only 1.0% of heterogeneity and was also not a significant moderator of effect ($F_{1,13}$=1.02, p=0.3302). Neither age ($F_{1,13}$=0.11, p=0.7436) nor amount of CFA injected ($F_{1,13}$=0.17, p=0.6848) accounted for any heterogeneity. All animals were tested less than one week post-CFA.

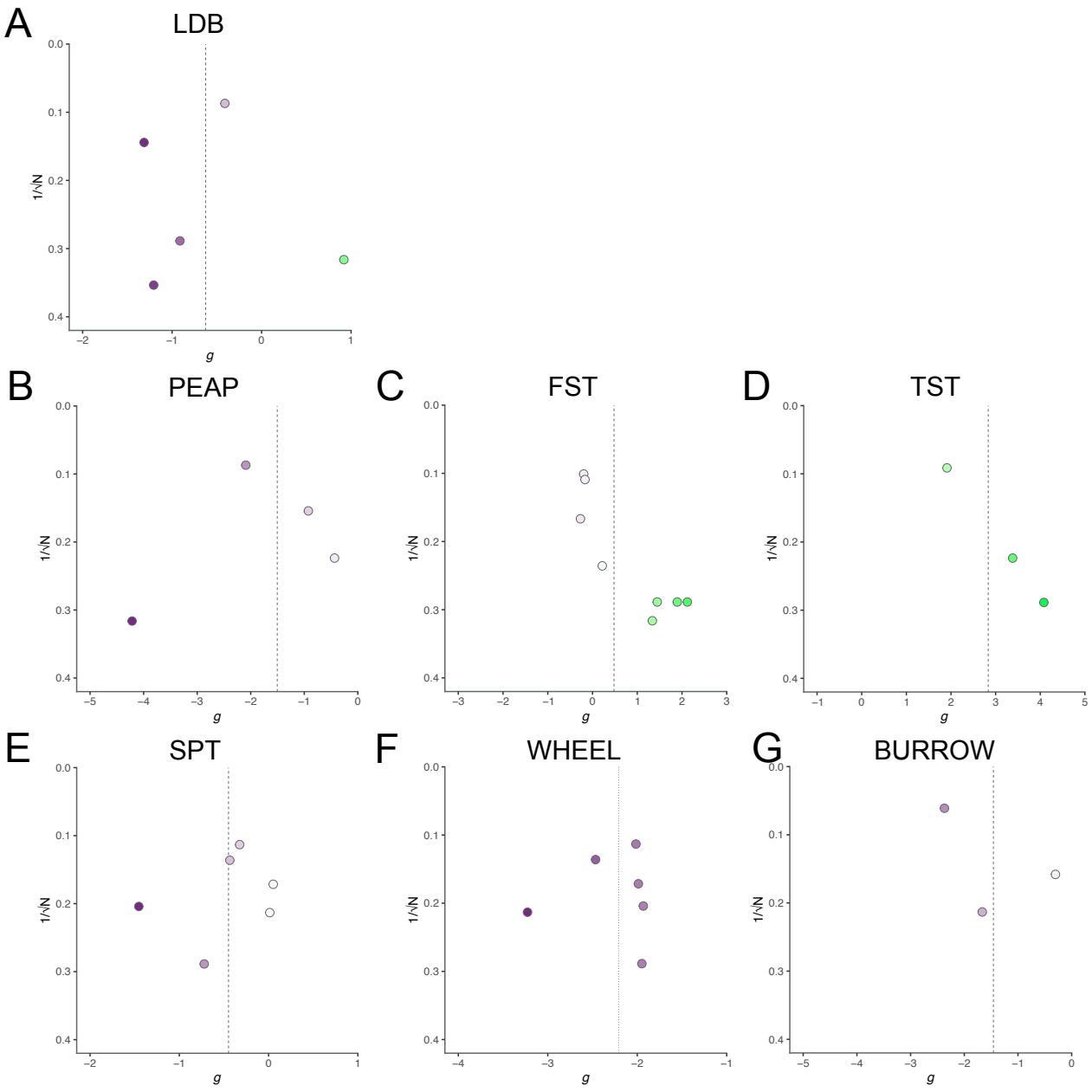


Figure S25. Funnel plots of "by experiment" effect sizes vs. $1/\sqrt{N}$ for behavioral assays with fewer than 10 studies, for which Egger's regression and trim-and-fill analysis are inappropriate. For "by experiment" statistics, please refer to corresponding forest plots in **Figs**. **S3**-**S12**.

	EPM	OFT	LDB	PEAP	FST	TST
Species	*			*		
Strain		**		*	*	
C57 Substrain		*	*			
Sex		*		**		
Animal Source	**	*	*		*	
Housing						
CFA Source	**					
CFA Injections						
Paw						
Naïve to Other				*		
Lighting						
Other	**					
Age	***					
ug CFA						
Duration					*	
Weeks						
Days						
Saline N						

Table S3. Summary of significant sub-group differences and meta-regressions for exploratory behavior and stress coping tests.

	SPT 1ST	SPT GE	WHEEL 1ST	WHEEL GE	BURROW 1ST	BURROW GE
Species					***	***
Strain					***	***
C57 Substrain						
Sex			*	*	**	**
Animal Source				*	***	***
Housing						
CFA Source						
CFA Injections						
Paw	*				***	***
Naïve to Other						
Lighting				*	**	**
Other					***	***
Age					**	
ug CFA					*	*
Duration		*			**	**
Weeks						
Days						
Saline N						

Table S5. Summary of significant sub-group differences and meta-regressions for natural rewards.

Study	Behavior	Intervention	Type	Result
Boyce-Rustay2010	PEAP	celecoxib, diclofenac duloxetine fluoxetine scopolamine	NSAIDs SNRI antidepressant SSRI antidepressant muscarinic antagonist	↑time on black/stim side ↑ time on black/stim side no effect no effect
Chen2012	PEAP	L-alpha-aminoadipate in anterior cingulate	astroglial toxin	↑ time on black/stim side
Cobos2012	wheel	naproxen, ibuprofen, diclofenac, celecoxib prednisolone morphine	NSAIDs glucocorticoid mu/delta opioid agonist	↑ distance traveled ↑ distance traveled ↑ distance traveled
Fang2020	splash TST SPT	baicalin	$GABA_AR$ PAM	↑ time grooming ↓ time immobile ↑ % sucrose preference
Gould2016	burrow	celecoxib, ibuprofen indomethacin morphine tramadol anti-NGF gabapentin diazepam	NSAIDs NSAID mu/delta opioid agonist mu opioid agonist; SNRI antibody calcium channel subunit regulator benzodiazepine	↑ grams burrowed no effect no effect no effect ↑ grams burrowed no effect no effect
Guan2020	OFT EPM	polydatin	glucoside/resveratrol derivative	↑ time in center ↑ time in open arms
Guo2016	OFT EPM	sesamin	lignan	↑ time in center ↑ time in open arms
Guo2018	OFT	CDPPB in anterior cingulate	mGluR5 PAM	↑ time in center
Hamann2016	FST	*V. megapotamica* morphine	phenolic aids, flavonoids mu/delta opioid agonist	↓ time immobile ↓ time immobile
Jin2020	OFT EPM	inhibiting glutamatergic in layer 5 somatosensory to GABAergic in caudal dorsolateral striatum projection	optogenetic & chemogenetic	↑ time in center ↑ time in open arms
Laumet2020	FST	$Rag2^{-/-}$	devoid of adaptive immune T cells	prolonged ↑ time immobile
Le2014	FST	CX546, CX516	AMPAkines	↓ time spent immobile
Luo2020	OFT EPM	scopoletin	coumarin	↑ time in center ↑ time in open arms
Maciel2013	TST	imipramine fluoxetine bupropion dexamethasone indomethacin celecoxib dipyrone pregabalin	tricyclic SSRI NRI glucocorticoid NSAID NSAID prostaglandin inhibitor calcium channel subunit regulator	↓ time immobile ↓ time immobile ↓ time immobile no effect no effect ↓ time immobile ↓ time immobile no effect
	FST	imipramine bupropion + celecoxib	tricyclic NRI + NSAID	↓ time immobile ↓ time immobile
Negus2015	nesting	ketoprofen morphine U69,593	NSAID mu/delta opioid agonist kappa opioid agonist	↑ zones cleared ↑ zones cleared ↓ zones cleared
Omorogbe2018	LDB TST	Jobelyn	dietary supplement from grain	↑ time in light ↓ time immobile
Parent2012	EPM	morphine diazepam	mu/delta opioid agonist benzodiazepine	↑ time in open arms ↑ time in open arms
Qi2014	EPM	U0126 in PFC	MAPK inhibitor	no effect
Reker2020	OPTA	meloxicam	NSAID	↓ time in 40° reward zone
Stickney2021	wheel	morphine implant	mu/delta opioid agonist	withdrawal ↓ running more in CFA
Sun2016	OFT EPM	gastrodin	glucoside	↑ time in center ↑ time in open arms
Sun2020	OFT EPM	8-O-acetyl shanzhiside methylester	iridoid glycoside	↑ time in center ↑ time in open arms
Tian2017	OFT EPM	α-asarone	phenylpropanoid	↑ time in center ↑ time in open arms
Uhelski2012	PEAP	somatosensory lesion, hind-paw region somatosensory lesion, barrel cortex region	lesion	↑ time in light no effect
Wang2015	OFT EPM	ZBD-2	translocator protein ligand	↑ time in center ↑ time in open arms
Yue2018	OFT EPM	CPEB1-shRNA	RNA interference	↑ time in center ↑ time in open arms

Table S6. Summary of effects of pharmacological and experimental interventions on CFA-induced behavior.