Estimating information flow through a memory system:

the utility of meta-analytic methods for genetics

Tugce Yildizoglu[1,2], Jan-Marek Weislogel[1,2], Farhan Mohammad[1,2], Edwin S.-Y. Chan[3,4], Pryseley N. Assam[3,4] and Adam Claridge-Chang[1,2,5,*]

1. Program in Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School, Singapore 138673

2. Institute for Molecular and Cell Biology, Singapore 138673

3. Singapore Clinical Research Institute, Singapore 138669

4. Centre for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore 169857

5. Department of Physiology, National University of Singapore, Singapore 138673

* Corresponding author. Program in Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School, 61 Biopolis Drive, Singapore 138673. claridge-chang.adam@duke-nus.edu.sg

ABSTRACT

Critics of significance testing claim that this statistical framework promotes discrepancies by using arbitrary thresholds (α) to impose reject/accept dichotomies on continuous data, which is not reflective of the biological reality of quantitative phenotypes. Here we explore this idea and evaluate an alternative approach, demonstrating the potential for meta-analysis and related estimation methods to resolve discordance generated by the use of traditional significance tests. We selected a set of behavioral studies proposing differing models of the physiological basis of *Drosophila* olfactory memory and used systematic review and meta-analysis approaches to define the true role of lobular specialization within the brain. The mainstream view is that each of the three lobes of the *Drosophila* mushroom body play specialized roles in short-term aversive olfactory memory [1-5], but a number of studies have made divergent conclusions based on their discordant experimental findings [6-8]. Multivariate meta-regression models revealed that short-term memory lobular specialization is not in fact supported by the data, and identified the cellular extent of a transgenic driver as the major predictor of its effect on short-term memory. Our findings demonstrate that meta-analysis, meta-regression, hierarchical models and estimation methods in general can be successfully harnessed to identify knowledge gaps, synthesize divergent results, accommodate heterogeneous experimental design and quantify genetic mechanisms.

LAY SUMMARY

Significance testing is a statistical method widely used to assess the effect of gene variation, particularly in genetic models e.g., mice and flies. While many traits vary continuously, significance testing is designed to produce a simple yes/no outcome that is unsuitable for smoothly varying effects. For decades, statistical texts have proposed that significance testing has a distorting effect. We examined the influence of significance testing in vinegar fly research on short-term olfactory memory, an extensively-studied quantitative phenotype in a major model genetic system. Significance tests have been previously used to show that a particular brain sub-region - the gamma lobe - is highly specialized for short-term memory. We re-analyzed published data using estimation methods that place emphasis on quantification: meta-analysis and hierarchical models. We show that the gamma lobe shares memory processing roles with two other brain lobes. Using neuronal cell count data, we also show an absence of even relative specialization, revealing that memory processing is distributed across neurons in all three lobes. We propose that significance testing distorts other areas of genetic analysis in similar ways; our data indicate that this can be ameliorated with the adoption of estimation statistics instead.

INTRODUCTION

Contradictory results from research are commonplace. Discordance stems from sampling error and methodological differences, both sources of variability that are largely unavoidable. One concern is the widespread acceptance of weak significance testing power; a recent report revealed that many neuroscience studies have a test power below 40% [9], and the same pattern is likely to be seen across other disciplines. However, critics of significance testing itself claim that this statistical framework needlessly accentuates differences. The numerous conceptual and practical limitations of significance tests [10] include the inherent volatility of p-values, even when there is moderate statistical power [11]. Simulation demonstrates that test results from different studies can easily be discordant due to sampling error alone, even assuming flawless methodological standardization [11]. Moreover, significance testing may exacerbate discordance by using an arbitrary threshold to force a binary outcome (reject/accept) from continuous data [12]. To illustrate, a pair of alpha 0.05 tests on two replicated experiments with identical effect sizes could produce p-values of 0.049 and 0.051: in the significance testing framework these results are starkly discordant, when in reality the biological outcome is all-but the same [12]. The arbitrary reject/accept dichotomy might also lead to the false impression that a substantial (but non-statistically significant) effect is irrelevant. Conversely, a highly powered sample size could give the misleading impression that a minuscule (but statistically significant) effect is of great importance [11]. Thus, some consider that fields relying solely on significance testing to draw their conclusions are particularly susceptible to discrepancies and may be incapable of resolving apparent irreproducibility.

In medical research, the complementary methods of systematic review and meta-analysis are routinely used to synthesize evidence from multiple studies and to reconcile divergent findings [13]. However, such approaches are rarely applied to basic research fields like neuroscience. Taking a mainstream sub-field of neuroscience as an example, a PubMed search in late 2014 with the phrase "meta-analysis AND (learning OR memory) AND mouse" identified fewer than ten studies in a field of >35,000 articles. We therefore decided to ask whether meta-analytic methods could be used to evaluate the possible influence of significance testing dichotomization. In seeking a suitable research field we required an unresolved hypothesis for which the published studies included adequate sample sizes for meaningful analysis, and used a standardized protocol so that the data would not be dominated by sampling error (weak statistical power) and methodological heterogeneity. We selected the investigation of the neuronal mechanisms of olfactory memory in *Drosophila melanogaster*. Olfactory memory in *Drosophila* is measured using the classical T-maze olfactory conditioning assay, where groups of flies are conditioned by pairing an odor with an electric shock and subsequently assessed for their ability to avoid the conditioned odor when given a choice of two different odors presented at the end of the maze arms. A particular strength of the T-maze is its use of hundreds or thousands of animals in a single

experiment [14], which helps to minimize the sampling error that is often inherent in rodent assays and assays using smaller numbers of insects [15,16]. In addition, both the T-maze apparatus and the training regime is largely standardized between labs [14].

Thirty years of T-maze experiments have elucidated many of the genetic, molecular and neural mechanisms of olfactory learning [1-5,17]. A landmark study showed that restoring the adenylyl cyclase gene *rutabaga (rut*) to a brain structure called the mushroom body is sufficient for short-term olfactory memory [6], connecting memory formation to cyclic adenosine monophosphate-mediated plasticity [18]. Experiments using inhibition of synaptic transmission by temperature-sensitive *shibire (shi)* [19-21] showed that neurotransmission from the mushroom body is essential [20,22]. Targeted expression of genes in specific neuronal circuits is possible with the use of transgenic 'driver' lines [23]. Manipulations based on *rut* restoration and *shi* inactivation form the foundation of a large number of studies aiming to further define the role of the mushroom body in olfactory learning. The mushroom body itself exists as three anatomically distinct lobes, αβ, α′β′, and γ [24]; studies on middle- and long-term memory (MTM and LTM) have revealed distinct lobe requirements in the different memory phases [21,25-28]. However, the three lobes' specializations remain unclear when it comes to short-term memory (STM). While the mainstream view is that *rut* activity in the γ lobes is sufficient to rescue STM [8], some studies have alternately concluded that *rut* restoration can only partially rescue [7], or is merely of importance to STM [6]. There is similar controversy on the role of *rut* activity in the αβ lobes, with *rut* restoration said to have either no effect [8], or to partially rescue STM for certain odors [7].

In the present study, we aimed to evaluate the mainstream view that there is strong lobular specialization of STM function in the mushroom body, and to assess the extent to which the varying perspectives on this subject resulted from significance testing's forced dichotomization. Using meta-analytic methods, we examined the proposals that restoration of *rut f*unction to the γ lobes alone is sufficient to rescue wild type STM and that only *shi* function in the γ lobes is necessary for STM. In both cases, meta-analysis of published studies spanning more than a decade found no evidence for strong lobular specialization. A subsequent analysis with multi-level meta-regression, an advanced estimation technique, revealed that numbers of mushroom body cells explained nearly all transgenic effects. These results confirm claims made by statistical texts that systematic review, meta-analysis and related estimation methods can be applied to resolve currently conflicting data and give new quantitative perspectives. In addition to its role in review, we conclude that routine use of both basic and advanced estimation methods would aid the planning, analysis and interpretation of research.

RESULTS

Systematic literature review of *rutabaga* and *shibire* interventions in short-term aversive olfactory memory

The review yielded ten studies that fulfilled the criteria (Figure 1A). Seven studies contained 81 experiments related to *rutabaga* restoration [6-8,22,29-31], with a total of 748 experimental iterations and 745 control iterations (see Table 1). Each iteration is the mean of two half-PI scores, which typically each use 50-100 flies, thus representing an estimated total of 150,000-300,000 assayed flies. Table 1 also lists the 5 studies that contained 37 experiments related to *shibire*-mediated inactivation [7,20,21,25,29], 263 experimental iterations and 265 control iterations, giving a total of 50,000-100,000 flies.

Experimental variability

Despite standardization of aspects of the T-maze, some methodological variation between studies was observed, including different control genotypes, varying odor pairs, temperatures, shock voltages, humidity and post-training delay times prior to testing (Table 1). These differences, along with other uncontrolled variables common to behavioral experiments, would explain the variability seen in data from control experiments (Figure 1B). We found considerable heterogeneity in several of the meta-analyses. In the six *rut* analyses, overall heterogeneity was low in three ($I^2 < 50\%$), and high in three ($I^2 > 75\%$); subgroup heterogeneity (i.e. variance due to genotype differences) was low in four, and high in two (Figures 6-11). In the *shi* analyses, overall heterogeneity was high in two and moderate in one, while their subgroup heterogeneity values were 34%, 64% and 80%.

Rutabaga function is required for 60% of wild type learning

We aimed to estimate the learning contribution made by restoring *rutabaga* function to each of the three lobes. The meta-analyses on *rutabaga* experiments produced 6 meta-analytical estimates of the effects of manipulating *rut* in the mushroom body lobes (Figure 2B). Data pooled from rut^1 and rut^{2080} reveal that the strong *rut* hypomorphic alleles reduce learning to 40% of wild type (-60% [95CI -56, -64]) (Figure 2B, 6). The forest plot in Figure 2A illustrates the individual effect sizes from 36 experiments and pooled effect sizes of the *rut* alleles (complete forest plot is shown in Figure 6). The data exhibit substantial overall heterogeneity ($I^2 = 76\%$) and genotype subgroup heterogeneity ($I^2 = 88\%$). This heterogeneity may derive from the methodological variation noted above, but in the case of the strong *rut* alleles we note that the weakest effect is seen in the rut^{2080}; *UAS-rut* subgroup (-45% [95CI -38, -52]), suggesting leaky expression from the transgene as one possible source (i.e. expression from the *UAS-rut* transgene independent of GAL4 transcriptional activation).

Rutabaga restoration to the γ lobes rescues 26% of wild type STM

Some studies have reported that complete rescue requires *rut* restoration in both αβ and γ lobes [7], while others report that restoring *rut* activity in the γ lobe is sufficient to rescue STM, and that the αβ lobes' *rut* activity has little or no STM role [8]. We used the meta-analytic data to specifically examine the lobular specialization hypothesis (Figures 5-10). The overall *rutabaga* loss-of-function effect was used as a reference point to which we compared the lobe restorations, shown in Figure 2B. Restoring *rut* function to each of the lobes revealed partial rescue: α′β′ rescues by 6% [95CI -1.5, 13.5], αβ rescues by 12% [95CI 2, 22] and γ rescues by 26% [95CI 17, 35]. When *rutabaga* was restored to both the αβ and the γ lobes, memory was rescued by 52% [95CI 50, 55]. Restoring *rutabaga* to all three lobes gave only 1% additional improvement (53% [95CI 47, 59]) compared to the rescue in the αβ + γ lobes, therefore *rut* in the α′β′ cells appears to have a minor effect on STM. Of the enhancer trap drivers included in the γ meta-analysis, 201Y contains a minority of αβ cells [32]. A variant analysis that removed 201Y from the γ group and reassigned it to the αβ + γ group resulted in weaker effects for both: only 20% [95CI 10, 31] γ rescue, while αβ + γ rescue was reduced to 49% [95CI 46, 52]. Taken together, these results are incompatible with the hypothesis that restoring *rut* activity to the γ lobe alone is sufficient to rescue the *rut*⁻ phenotype. From the lobe perspective, we conclude that normal STM requires *rut* function in both αβ + γ lobes.

Heating flies above 30°C impairs short-term memory

Using the temperature-sensitive alleles of *shibire* to block neurotransmission requires heating flies to over 30°C, which can lead to additional heat-related effects [28]. Researchers accommodate this possibility with separate 'heat control' flies that do not express *shi^{ts}*. We estimated the magnitude of this effect by meta-analysis, shown in Figure 3A (complete forest plot in Figure 11). Data pooled from 23 such experiments with three types of genotype (wild type, *Driver-GAL4/+* and *UAS-shi^{ts}/+*) revealed that the overall effect of heating flies from the permissive temperature (20-26°C) to 30-35°C is a 17% [95CI 12, 22] reduction in memory. This decrement can be expected to affect the *UAS-shi^{ts}* inactivation data from the same studies, so we used 83% of wild type memory in Figure 3B as the zero reference point to estimate the specific effects of lobe inactivation.

Neurotransmission from the αβ + γ lobes accounts for 61% of STM

Drivers that express in both the αβ and γ lobes reduced performance by 61% [95CI 50, 72] relative to heated control flies. Inactivating the αβ lobes produced a 25% [95CI 14, 37] reduction in STM. The best estimate for γ lobe inactivation is a 6% reduction [95CI 35% reduction, 24% increase] relative to heated controls. This γ lobe estimate appears to be negligible, but has very wide confidence intervals and is drawn from only a single experiment with three iterations. Surprisingly, the literature review found no <5 min STM data on the impact of *shibire^{ts}* inactivation of either the entire mushroom body (All lobes) or the α′β′ lobes (empty columns in Figure 3B); at the time of the review the only studies reporting results for these interventions

examined later memory, at 15 min or beyond [28]. The substantial decrement in the αβ lobe inactivation experiments (25% reduction) is incompatible with the idea that this lobe plays only a negligible role in STM. The paucity of data for γ, α′β′ and All lobes in STM highlights an area that would benefit from future experimental attention.

Cell number accounts for the majority of driver variation

Observing high heterogeneity (I^2) in some of the meta-analyses, we attempted to identify the source of variability, and examine the original hypothesis from a different perspective. Electrophysiological evidence [33] and anatomical connectivity analysis [34] indicate that the Kenyon cells, the intrinsic neurons of the mushroom body, are randomly connected to their olfactory input neurons. The lack of structured connectivity suggests that, for some or all odor-related functions, individual Kenyon cells are interchangeable; thus raising the possibility that a cell's lobular identity might be less important than its participation in a stochastically nominated odor-responsive ensemble. As three of the seven relevant meta-analyses showed driver heterogeneity as accounting for more than half of their variance, we asked whether the number of cells captured by a driver could explain some of the unaccounted variance. We extracted cell count data from an anatomical study that counted Kenyon cells for many of the drivers [32]. The driver-specific meta-analytic STM estimates were subjected to an initial simple linear regression against the drivers' available cell counts in both *rut* restoration and *shi*[ts] inactivation. These indicated that cell numbers accounted for about 80% of the driver memory variance (*rut* R^2 = 0.79 [95CI 0.39, 0.94], p=2.5 x 10^{-4}; *shi*[ts] R^2 = 0.77 [95CI 0.14, 0.96], p=8.4 x 10^{-3}). As simple linear regression is unable to account for the full complexity of such hierarchical data, we constructed hierarchical, multivariate, weighted meta-regression models accommodating other variables that might explain some of the variance induced by differences in experimental design. These models were also able to account for the clustering of experiments within studies and for the shared control design in *rut* experiments, and included weighted estimates for each driver by the number of contributing experiments (described fully in Methods). The hierarchical meta-regression model of *rut* showed a strong relationship with driver cell count, generalized-R^2 = 0.84 [95CI 0.79, 0.89] (Figure 4A). The meta-regression model of *shi* data similarly revealed a large effect size for the cell count relationship, generalized-R^2 = 0.88 [95CI 0.84, 0.92] (Figure 4B). Compared with simple linear regression, the hierarchical models revealed stronger trends with substantially improved precision. These results are incompatible with the strong lobular specialization hypothesis of *rut* and *shi* function. Rather, drawing on data from thousands of T-maze iterations (N = 1008, 1006) while accounting for experimental heterogeneity, they constitute compelling evidence that each driver's extent of neuronal expression can account for the majority of that driver's short-term memory effect.

Kenyon cells in different lobes make equivalent contributions to STM

Different Kenyon cell drivers' varying impact on learning is primarily a result of how many cells they are expressed in: cell count as the overwhelmingly dominant factor therefore excludes highly specialized roles for *rut* and *shi* in different lobes' Kenyon cells. However, it is possible that minor quantitative differences explain the remaining unaccounted for 12-16 % of STM variance in the meta-regression models. Within the overall memory-cell count trend in Figure 4A, several drivers' estimates do not fall on the regression line. To account for such deviations from the overall cell number trend, we aimed to factor out cell number and focus specifically on the potency of each neuron captured by a driver. We built new models in which the learning effect size of each driver line was first divided by the number of expressing cells, and weighted hierarchical meta-regression models were then used to perform synthesis by lobular category. These models produced estimates of a typical Kenyon cell's effectiveness within each lobe category (Figure 4C & D). The *rut* rescue-per-cell data and the *shi* loss-per-cell data both show that there are no substantial differences between any lobe categories. In summary, when cell numbers are taken into account, the evidence does not support the strong lobular specialization hypothesis. Instead, it shows that lobular *rut* function is non-specialized and that STM makes use of all available functioning Kenyon cells.

DISCUSSION

Previous studies using significance testing concluded that differences between mushroom body lobes exist that reflect functional specializations in the various memory phases (STM, MTM and LTM). These conclusions about lobular specialization included the idea that γ lobe *rut* function is sufficient for STM formation. The aim of the present study was to specifically examine the strong lobular specialization STM hypothesis. Surprisingly, the synthetic evidence is incompatible with lobular specialization, and supports the alternative idea that STM function is generalized across lobes.

Meta-analysis of strong *rut* hypomorphic alleles confirmed that they cause a 60% reduction in STM. As previously reported in the literature, the other 40% must be mediated by other molecular factors either in the Kenyon cells or elsewhere. Restoring *rut* activity with lobe-targeting drivers revealed that partial rescue occurs in both the γ and αβ lobes (mean 26% and 12%), with a partial rescue even in the α′β′ lobes (mean 6%). To rescue the majority of lost function, *rut* had to be expressed in both αβ and γ lobes (Figure 2B). These data are incompatible with the hypothesis that the lobes' *rut* activity in the γ lobe is absolutely or strongly specialized for STM. With the synthesized evidence failing to support strong lobular specialization of *rut* in STM (Figure 2B), we considered an alternative hypothesis: that cell extent is the main predictor of a transgenic driver's STM impact. Indeed, multivariate meta-regression models incorporating cell count show that the dominant factor influencing STM is the number of Kenyon cells targeted by a specific driver line, for both *rut* and *shi* effects (Figure 4A, B). This result refutes the hypothesis that the mushroom lobes are specialized for aversive STM function. Rather, the linear relationships lead us to conclude that the different lobes' cells have similar potency for STM with regard to *rut* and *shi*-dependent memory processes.

Despite the paucity of experiments for *shi* in the γ, α′β′ and All lobes categories, the available data were sufficient to allow construction of a precise model of the relationship between driver cell count and memory. If STM relied on neurotransmission from a highly inter-dependent Kenyon cell ensemble, we would anticipate that *shi*^{ts} inhibition of small subsets of these cells would have a large effect. Instead, the observed linear trend between driver cell count and STM impact (Figure 4B) supports a model in which *shi*-dependent memory function in the αβ and γ cells occurs autonomously in individual cells or small groups of cells. It appears that strong qualitative specialization of lobular neurotransmission emerges over the subsequent minutes and hours as later memory forms [26,28].

This investigation serves as a case study in how meta-analysis and related estimation methods can help animal behavior research specifically, and biological analysis in general. Recent commentary has focused attention on reproducibility [9,35,36] and replication [37]; both of these issues are in part connected to significance testing. An encouraging aspect that was revealed as a part of this study is that the existing published data could support precise estimation with hierarchical modeling, suggesting firm data integrity. Significance testing remains the dominant statistical methodology in neuroscience and many other life sciences [38], despite its tendency to amplify interpretive variation with arbitrary thresholds. For this and other reasons, significance testing has been controversial in the behavioral sciences for half a century [39],

but alternatives have yet to be implemented and adopted by the field. Estimation is an alternative data analysis framework that places the emphasis on effect sizes and the meta-analytic perspective. Exhortations in editorials and textbooks for researchers to favor estimation statistics for data analysis [10,11,40,41] have so far had little effect: neuroscience, biology, psychology and economics remain predominantly significance testing fields. This study shows how systematic review in conjunction with several meta-analytic techniques enable the synthesis of relevant available evidence so as to address inconsistencies in a field and reveal unexpected patterns in published data. Estimation as a fundamental statistical framework is also suitable for use in primary research; modern statistical texts advise that reporting effect sizes with their confidence intervals, along with the use of graphical methods, are the rightful priorities of primary data analysis [11,12,40]. Hierarchical models can similarly be applied routinely to analyze primary data with complex experimental designs, such as behavior experiments conducted in different labs [42] or by differing protocols within a lab [43,44], replacing basic methods such as ANOVA. The methods demonstrated above represent a superior statistical framework for all phases of biological research: planning, analysis, interpretation and review.

MATERIALS AND METHODS

Eligibility criteria and information sources

All information was sourced with searches of PubMed. To be eligible for consideration for inclusion in the systematic review each study was required to meet the following criteria: containing olfactory STM experiments on *Drosophila melanogaster* using the classic T-maze apparatus and a single training cycle [14]; reporting of the relevant control and experimental data as a Performance Index (PI); detailing the relevant genotypes and the number of experimental iterations (N or sample size). In addition, as STM is thought to begin to transition to MTM shortly after training [17], we defined STM as using a post-training delay of 5 minutes or less. All studies selected contained transgenic manipulations of the Kenyon cells targeted to one or more of the 3 lobes (αβ, α′β′, and γ). For the systematic review of *rut* function in the Kenyon cells, studies included use of a hypomorphic allele of the *rut* gene, transgenic drivers and *UAS-rut* expression constructs. Experiments using temporally controlled expression of *rut* were excluded to eliminate the possibility of heterogeneity associated with incomplete restoration due to variations in expression longevity or strength. For the systematic review of endocytosis-dependent neurotransmission in the Kenyon cells, studies included a *UAS-shi^{ts}* transgene in combination with transgenic drivers and heat treatment. Experiments that shifted *shi^{ts}* flies to different temperatures between training and testing were excluded to eliminate the possibility of heterogeneity due to these manipulations; only experiments using the conventional permissive-restrictive (cool-warm) comparison were included. Following the lead of the great majority of the STM literature, we did not attempt to analyze the acquisition, storage and retrieval phases of STM.

Database search

The systematic literature search was conducted as follows and is shown as a diagram in Figure 1A. On the 11th July 2013, the search phrase ((((Drosophila) AND (learning OR memory)) AND (mushroom OR Kenyon)) AND ("2000"[Date - Publication] : "3000"[Date - Publication]) NOT review[Publication Type] was used to query PubMed, and the resulting 279 records were downloaded as two .nbib files. These files were imported into Papers2 software, and then exported as EndNote .xml. This file was loaded into EndNote X4, copied into Excel, and then imported into Apple Numbers with all bibliographic information including Title and Abstract stored in one row per record. This was then used to screen the records' titles, abstracts and was also used to record the results of the full text screen and the detailed experimental design screen.

Study selection

We designed the literature selection process to identify experiments that examined aversive olfactory STM (testing five minutes or less after training) in *Drosophila* as observed in the classic T-maze apparatus. We further aimed to focus the analysis on the two kinds of experiments most commonly used to understand the role of the three mushroom body lobes and the mushroom body intrinsic neurons (Kenyon cells). The first type of experiments was the usage of transgenic *rutabaga* (*rut*) to restore adenylyl cyclase function to one or

more lobes in *rut* mutant flies; the second type included experiments targeting transgenic temperature-sensitive SHIBIRE (SHI[TS]) protein to the lobes to disable dynamin-dependent neurotransmission. The SHI[TS] proteins form part of the dynamin endocytosis complex and poison its function when flies are transferred to the restrictive temperature [45]. The exact odor pairs under investigation were explicitly disregarded in this analysis; rather, experiments containing the full variety odor pairs were included to enable us to arrive at the most general conclusion about mushroom body function.

Two investigators (TY and JMW) performed the literature review independently and discrepancies were resolved collaboratively with a third investigator (ACC). The 279 records yielded from the PubMed search were screened in four stages to systematically exclude studies: title review, abstract reading, full text scan and a detailed review of experimental design. This process is described in Figure 1A; we used title and abstract information to discover a set of *Drosophila* behavioral studies that were likely to include aversive olfactory conditioning in adult fly (n = 65 studies) and then scanned these full text articles to find *rutabaga* restoration or *shibire^ts* experiments in the MB lobes. The final stage in the selection ("Experimental Design" in Figure 1) excluded three studies that did not meet the eligibility criteria listed above: one did not use or report an isogenic permissive control [46]; a second did not report sample sizes and used a post-training interval of 15 minutes [28], i.e. 10 minutes later than the original criterion and 12 minutes later than other studies included; a third used pharmacogenetic temporal control of *rut* restoration [47].

Data item extraction

Two investigators (TY and JMW) extracted data independently using the measuring tool in Adobe Acrobat Pro; any discrepancies between the two extractions were resolved collaboratively. The following data were collected from each of the included experiments: author, year of publication, figure and panel numbers, genotype, mean Performance Index (PI) [48] with corresponding SEMs and the number of experimental iterations (N) for each mean PI value for each intervention and its related control group. To calculate STM percentages we identified a non-intervention control for each experiment, using the control that was the most similar to the experimental animals. For the *rut* restorations the closest available controls ranged from otherwise isogenic *rut^+* siblings to generic wild type (e.g. Canton-S). For the *shi^ts* experiments, including the heat-effect experiments, we used the permissive temperature controls. We also extracted experimental conditions: time delay between training and testing, odor pair, temperature, voltage, current type and relative humidity. One study's *rut* restoration data were plotted with superimposed error bars, precluding their extraction and inclusion in the review [25].

Driver line classification

Driver lines were classified by lobe expression pattern according to the original studies themselves, except for the MB247 line which was thought to drive expression in all lobes [21], but is now characterized as primarily driving expression in the αβ and γ lobes [26,32]. In addition, while several studies used 201Y as a

Yildizoglu *et al.*

γ driver, there is more recent evidence that 201Y also drives in a minority of αβ cells [32]; we accommodated this by doing primary analysis counting 201Y as γ, but also doing a variation in which it was counted as αβ + γ.

Summary measures

For each experiment we calculated the intervention's effect as a percentage change relative to the control PI. All the meta-analyses were carried out for the percentage change metric as well as the raw change in PI; the results were equivalent. We chose to report data as percentage changes for easier interpretation. The histogram in Figure 1B shows that control PI scores vary considerably across experiments; using a percentage change re-scales the phenotypes to each experiment's wild type memory. A percentage not only reports how far a phenotype is from wild type memory but also sets a lower bound (0% memory). The standard error of each percentage change was calculated using the delta method approximation [49,50].

$$SE_{pooled} = \frac{mean_{experimental}}{mean_{control}} \sqrt{\left(\frac{SE_{experimental}}{mean_{experimental}}\right)^2 + \left(\frac{SE_{control}}{mean_{control}}\right)^2}$$

Synthesis of results

Review Manager software {ReviewManagerRevM:wk} was used to perform nine meta-analyses: six on the *rutabaga* data, three on the *shibire* data. One random effects model meta-analysis was carried out for each mushroom body lobe and any available combinations; within each meta-analysis a subgroup analysis was performed for each driver line, except for the *rut* mutant and heat effect controls analyses, where genotype subgroups were used. Table 1 gives full details, Review Manager file is provided as Supporting Dataset 1. Complete forest plots of the six *rut* and four *shi* meta-analyses are shown in Figures 5-13. No meta-analysis was possible for *rut* restoration to the γ lobes as only one published experiment was found. Subgroup analysis of the driver lines was pre-specified. The I^2 statistic was used as a measure of the percentage contribution of heterogeneity to the total variance in each meta-analysis, including subgroup heterogeneity [51]. For ease of interpretation, summary plots showed learning as a percentage of wild type learning; these were calculated by addition of the impairment effect size to 100%. We report p-values from a two-sample t-test with unequal group variances in the *rut* and *shi* summary plots, and from a t-distribution transformation for the cell count regression. Otherwise, percentage effect sizes and their 95% confidence intervals were used to interpret all results [11]. All 95% confidence intervals are given in the form: [95CI lower, upper].

Meta-regression approach

Driver cell count data were extracted from a single anatomical study [32]. Initial examination of the relationship was done with MATLAB's simple linear regression function (LinearModel.fit.m) on the mean

values. However, this method does not account for many important aspects of the data. To accommodate the complex nature of the data, we performed multivariate hierarchical weighted meta-regression analyses of the driver effects using generalized linear mixed models (GLMM) in SAS version 9.3 software (SAS Institute, Cary, North Carolina; PROC GLIMMIX). For experiment k with appropriate control group j in study i, the outcome PI_{ijk} (raw change or relative percentage change) was modeled using GLMM taking into account the following:

- *The meta-analytic nature of the data:* each PI_{ijk} was estimated with a certain level of precision in the primary study/experiment. PI_{ijk} were weighted in the GLMM by their corresponding precision or inverse variance $\left(1/Var(P_{ijk})\right)$ with more weight assigned to more precise PI_{ijk}, as in the meta-analyses.

- *Relevant experimental design factors* (X_{ik}) were corrected for in the GLMM to reduce the variance induced by differences in design factors between individual experiments and studies. Univariate and multivariate GLMM models were developed by including one and more-than-one design factors as independent variables in the GLMM respectively.

- *Clustering*: multiple experiments are clustered (nested) within each study and this clustering may introduce extra variability or dependence due to laboratory and personnel preferences (practice) in conducting experiments. Studies were modeled as clusters (b_i) through a random effect with variance τ.

- *Shared Controls*: *rut* restorations within experiments were calculated based on a shared control, which created dependencies (correlation) between *rut* restoration effects that shared control groups. Therefore residuals (ε_{ijk}) based on the same (shared) controls were correlated and residuals based on different controls were independent. Due to convergence issues arising from a paucity of data we assumed a constant correlation (ρ) between residuals based on the same shared controls and modeled the residual variance-covariance matrix (Σ) with a block compound symmetry structure – blocked by shared controls, leading to conditionally independent residuals. A simple constant-variance diagonal variance-covariance matrix was used for the *shi* experiments, as matched controls were available, leading to independent residuals.

Coupling all these aspects together yielded the following univariate and multivariate weighted GLMM:

$$PI_{ijk} = \sum_{i,j} \beta_{ij} X_{ij} + b_i + \varepsilon_{ijk},$$

$$b_i \sim N(0, \tau^2),$$

$$\varepsilon_{ijk} \sim N(0, \Sigma), \text{ where } Corr(\varepsilon_{ijk}, \varepsilon_{ijk'}) \neq \rho, \ Corr(\varepsilon_{ijk}, \varepsilon_{ij'k}) = 0, \text{ and } Corr(\varepsilon_{ijk}, \varepsilon_{i'jk}) = 0.$$

Construction of models

Model construction started with inspection of all the available independent variables based on univariate GLMM. From Table 1, these variables included which pair of odors was used ('ODOR PAIR'), experimental temperature ('TEMPERATURE'), delay time between testing and training ('TIME'), shock voltage ('VOLTAGE'), voltage type ('AC/DC') and relative humidity ('RH'). The data are provided in Supporting Dataset 2. It was noted that the ODOR PAIR variable consisted of numerous categories, which would dramatically increase the degrees of freedom, so we considered replacing this with an approximation of the variable instead. Since benzaldehyde is known to stimulate gustatory receptors as well as olfactory receptors (and thus might have a different dependency on mushroom body function from other odorants), we used the presence or absence of benzaldehyde ('BENZALDEHYDE') as a proxy for ODOR PAIR. Of these variables, RH, AC/DC and VOLTAGE were both censored in a large proportion of experiments, and (for non-censored experiments) had mainly trivial and non-statistical effects on learning; these variables were excluded from subsequent models. TIME and BENZALDEHYDE data were available for all experiments. For *rut* experiments, both variables showed substantial and statistical influences on learning (TIME generalized-R^2 = 0.26 [95CI 0.15, 0.36]; BENZALDEHYDE generalized-R^2 = 0.28 [95CI 0.17, 0.39]), so these were incorporated into further multivariate meta-regression models. For the *shi* experiments, only TIME had a substantial influence on learning outcome (TIME generalized-R^2 = 0.12 [95CI 0.04, 0.21]). Multivariate GLMM were used to account for and extract the effect of the relevant independent variables by obtaining residuals from the respective multivariate GLMM. We calculated a residual learning effect by summarizing the residuals by drivers and rescaling them by subtracting the wild type memory reference value (*shi* = 83%; *rut* = 40%). The residual learning effect was regressed against cell counts in a linear meta-regression that was weighted by sample size (the number of experiments contributing to each driver). The learning-per-cell model was built by first dividing each driver's effect (and standard error) by its cell counts, and then fitting a multivariate GLMM with lobe categories as the main independent variable, while adjusting for other relevant experimental design factors.

ACKNOWLEDGMENTS

JMW, TY and ACC did the systematic review; JMR and TY performed the data extraction, TY performed the meta-analyses; FM performed the linear regression analysis; PNA built the meta-regression models; ACC, PNA and ESYC guided the project; ACC wrote the manuscript with contributions from the other authors.

REFERENCES

1.   Keene AC, Waddell S (2007) Drosophila olfactory memory: single genes to complex neural circuits. Nat Rev Neurosci 8: 341–354. doi:10.1038/nrn2098.

2.   Busto GU, Cervantes-Sandoval I, Davis RL (2010) Olfactory learning in Drosophila. Physiology (Bethesda, Md) 25: 338–346. doi:10.1152/physiol.00026.2010.

3.   Kahsai L, Zars T (2011) Learning and memory in Drosophila: behavior, genetics, and neural systems. Int Rev Neurobiol 99: 139–167. doi:10.1016/B978-0-12-387003-2.00006-9.

4.   Davis RL (2011) Traces of Drosophila memory. Neuron 70: 8–19. doi:10.1016/j.neuron.2011.03.012.

5.   Perisse E, Burke C, Huetteroth W, Waddell S (2013) Shocking revelations and saccharin sweetness in the study of Drosophila olfactory memory. Curr Biol 23: R752–R763. doi:10.1016/j.cub.2013.07.060.

6.   Zars T, Fischer M, Schulz R, Heisenberg M (2000) Localization of a short-term memory in Drosophila. Science 288: 672–675.

7.   Akalal D-BG, Wilson CF, Zong L, Tanaka NK, Ito K, et al. (2006) Roles for Drosophila mushroom body neurons in olfactory learning and memory. Learning & Memory 13: 659–668. doi:10.1101/lm.221206.

8.   Blum AL, Li W, Cressy M, Dubnau J (2009) Short- and long-term memory in Drosophila require cAMP signaling in distinct neuron types. Curr Biol 19: 1341–1350. doi:10.1016/j.cub.2009.07.016.

9.   Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, et al. (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14: 365–376. doi:10.1038/nrn3475.

10.   Cohen J (1994) The earth is round (p < .05). American Psychologist 49: 997–1004.

11.   Cumming G (2012) Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. 1 pp.

12.   Ellis PD (2010) The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge University Press. 1 pp.

13.   Borenstein M, Hedges LV, Higgins J, Rothstein HR (2011) Introduction to meta-analysis. John Wiley & Sons, Ltd.

14.   Tully T, Quinn W (1985) Classical conditioning and retention in normal and mutant Drosophila melanogaster. J Comp Physiol [A] 157: 263–277.

15.   Heisenberg M, Wolf R (1979) On the fine structure of yaw torque in visual flight orientation of Drosophila melanogaster. Journal of Comparative Physiology A 130: 113–130.

16.   Claridge-Chang A, Roorda RD, Vrontou E, Sjulson L, Li H, et al. (2009) Writing Memories with Light-Addressable Reinforcement Circuitry. Cell 139: 405–415. doi:10.1016/j.cell.2009.08.034.

17.   Heisenberg M (2003) Mushroom body memoir: from maps to models. Nat Rev Neurosci 4: 266–275. doi:10.1038/nrn1074.

18.   Levin LR, Han PL, Hwang PM, Feinstein PG, Davis RL, et al. (1992) The Drosophila learning and memory gene rutabaga encodes a Ca2+/Calmodulin-responsive adenylyl cyclase. Cell 68: 479–489.

19. Kitamoto T (2001) Conditional modification of behavior in Drosophila by targeted expression of a temperature-sensitive shibire allele in defined neurons. J Neurobiol 47: 81–92.

20. Dubnau J, Grady L, Kitamoto T, Tully T (2001) Disruption of neurotransmission in Drosophila mushroom body blocks retrieval but not acquisition of memory. Nature 411: 476–480. doi:10.1038/35078077.

21. McGuire SE, Le PT, Davis RL (2001) The role of Drosophila mushroom body signaling in olfactory memory. Science 293: 1330–1333. doi:10.1126/science.1062622.

22. McGuire SE, Le PT, Osborn AJ, Matsumoto K, Davis RL (2003) Spatiotemporal rescue of memory dysfunction in Drosophila. Science 302: 1765–1768. doi:10.1126/science.1089035.

23. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118: 401–415.

24. Crittenden JR, Skoulakis EM, Han KA, Kalderon D, Davis RL (1998) Tripartite mushroom body architecture revealed by antigenic markers. Learn Mem 5: 38–51.

25. Schwaerzel M, Heisenberg M, Zars T (2002) Extinction antagonizes olfactory memory at the subcellular level. Neuron 35: 951–960.

26. Krashes MJ, Keene AC, Leung B, Armstrong JD, Waddell S (2007) Sequential use of mushroom body neuron subsets during drosophila odor memory processing. Neuron 53: 103–115. doi:10.1016/j.neuron.2006.11.021.

27. Weislogel J-M, Bengtson CP, Muller MK, Hortzsch JN, Bujard M, et al. (2013) Requirement for nuclear calcium signaling in Drosophila long-term memory. Sci Signal 6: ra33. doi:10.1126/scisignal.2003598.

28. Cervantes-Sandoval I, Martin-Pena A, Berry JA, Davis RL (2013) System-like consolidation of olfactory memories in Drosophila. J Neurosci 33: 9846–9854. doi:10.1523/JNEUROSCI.0451-13.2013.

29. Schwaerzel M, Monastirioti M, Scholz H, Friggi-Grelin F, Birman S, et al. (2003) Dopamine and octopamine differentiate between aversive and appetitive olfactory memories in Drosophila. J Neurosci 23: 10495–10502.

30. Thum AS, Jenett A, Ito K, Heisenberg M, Tanimoto H (2007) Multiple memory traces for olfactory reward learning in Drosophila. J Neurosci 27: 11132–11138. doi:10.1523/JNEUROSCI.2712-07.2007.

31. Scheunemann L, Jost E, Richlitzki A, Day JP, Sebastian S, et al. (2012) Consolidated and labile odor memory are separately encoded within the Drosophila brain. J Neurosci 32: 17163–17171. doi:10.1523/JNEUROSCI.3286-12.2012.

32. Aso Y, Grübel K, Busch S, Friedrich AB, Siwanowicz I, et al. (2009) The mushroom body of adult Drosophila characterized by GAL4 drivers. J Neurogenet 23: 156–172. doi:10.1080/01677060802471718.

33. Murthy M, Fiete I, Laurent G (2008) Testing odor response stereotypy in the Drosophila mushroom body. Neuron 59: 1009–1023. doi:10.1016/j.neuron.2008.07.040.

34. Caron SJC, Ruta V, Abbott LF, Axel R (2013) Random convergence of olfactory inputs in the Drosophila mushroom body. Nature 497: 113–117. doi:10.1038/nature12063.

35. Reducing our irreproducibility (2013) Reducing our irreproducibility. Nature 496.

36. Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med 2: e124. doi:10.1371/journal.pmed.0020124.

37. Ioannidis JPA (2012) Why science is not necessarily self-correcting. Perspectives on Psychological Science 7: 645–654. Available: http://pps.sagepub.com/content/7/6/645.full.

38. Hentschke H, Stüttgen MC (2011) Computation of measures of effect size for neuroscience data sets. Eur J Neurosci 34: 1887–1894. doi:10.1111/j.1460-9568.2011.07902.x.

39. Morrison DE, Henkel RE, editors (1970) The Significance Test Controversy. Chicago: Transaction Publishers. 1 pp.

40. Altman DG, Machin D, Bryant TN, Gardner MJ, editors (2000) Statistics with confidence: confidence intervals and statistical guidelines. 2nd ed. BMJ Books.

41. Ziliak ST, McCloskey DN (2008) The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives. Ann Arbor: University of Michigan Press. 1 pp.

42. Crabbe JC, Wahlsten D, Dudek BC (1999) Genetics of mouse behavior: interactions with laboratory environment. Science 284: 1670–1672.

43. Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, et al. (2014) Olfactory exposure to males, including men, causes stress and related analgesia in rodents. Nat Methods 11: 629–632. doi:10.1038/nmeth.2935.

44. Richter SH, Garner JP, Würbel H (2009) Environmental standardization: cure or cause of poor reproducibility in animal experiments? Nat Methods 6: 257–261. doi:10.1038/nmeth.1312.

45. Narayanan R, Ramaswami M (2001) Endocytosis in Drosophila: progress, possibilities, prognostications. Exp Cell Res 271: 28–35. doi:10.1006/excr.2001.5370.

46. Thum AS, Knapek S, Rister J, Dierichs-Schmitt E, Heisenberg M, et al. (2006) Differential potencies of effector genes in adult Drosophila. J Comp Neurol 498: 194–203. doi:10.1002/cne.21022.

47. Mao Z, Roman G, Zong L, Davis RL (2004) Pharmacogenetic rescue in time and space of the rutabaga memory impairment by using Gene-Switch. Proc Natl Acad Sci USA 101: 198–203. doi:10.1073/pnas.0306128101.

48. Quinn W, Harris W, Benzer S (1974) Conditioned behavior in Drosophila melanogaster. Proc Natl Acad Sci U S A 71: 708–712.

49. Oehlert GW (1992) A Note on the Delta Method. The American Statistician 46: 27–29.

50. Cramer H (1946) Mathematical Models of Statistics. Princeton NJ: Princeton University Press.

51. Higgins J, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ: British Medical Journal 327: 557–560.

FIGURE AND TABLE LEGENDS

Figure 1. Review overview.

A. Flow chart of systematic literature review procedure. The literature was reviewed in a five stage process, starting with a PubMed search that yielded 279 articles, followed by four screens of increasing detail, reviewing the article title, abstract full text and experimental design. A total of ten articles, two of which included relevant data for both rutabaga and shibirets experiments, were used in the meta-analyses. **B.** Histogram of performance indices for all control experiments identified by the review.

Figure 2. Meta-analyses of *rutabaga* mutant lines and targeted transgenic restoration.

Short-term memory data are expressed as percentages. **A.** A summary forest plot of learning changes observed in 340 experiments with *rut* mutant lines, with subgroups showing the differences between the various *rut* alleles and strains. Learning is expressed as a percentage change relative to wild type. The red diamond on the bottom line indicates that the overall impairment in learning in the *rut* hypomorphs relative to wild type controls is -60% [95CI -56%, -64%]. The complete forest plot is given in Figure 5. **B.** Summary estimates from the *rut* mutant meta-analysis and five meta-analyses of lobular restoration experiments. Learning is displayed as a percentage of wild type learning. The markers indicate the proportion of learning relative to wild type expressed as a percentage; error bars are 95% confidence intervals. To the right of the markers are numbers for the amount of rescue (R =) relative to the rut hypomorphs. N(E) and N(C) are the experimental and control iterations respectively. Except for the $\alpha'\beta'$ lobes (p=0.17), all lobe categories showed a statistically significant partial rescue of learning ($\alpha\beta$ p=0.029, γ p<1 x 10-45, $\alpha\beta$+γ p=1.1 x 10-16, all lobes p<1 x 10-45) when compared with rut learning.

Figure 3. Meta-analyses of *shibire*[ts] inhibition of neurotransmission in the mushroom body lobes.

Learning data are expressed as percentages. **A.** A summary forest plot of learning changes in heat treatment controls, with subgroups showing the differences between 3 types of controls. Learning is expressed as a percentage change relative to wild type. The red diamond on the bottom line indicates that the overall impairment in learning in flies exposed to elevated temperature is -17% [95CI -12%, -22%]. A complete forest plot is shown in Figure 11. **B.** Summary estimates from the heat exposure controls and three meta-analyses of lobular inactivation experiments. Colored markers correspond to diamonds in panel A. Learning at the restrictive temperature is shown as a percentage of learning at the permissive temperature; error bars are 95% confidence intervals. To the right of the markers are numbers learning impairment (Δ* =) relative to the synthetic heat effect control. N(R) and N(P) are the restrictive and permissive iterations respectively. The $\alpha\beta$ lobes (p=0.0001) and the $\alpha\beta$+γ combination (p<1 x 10-45) show statistically significant impairment while the γ lobes do not (p=0.7071). The γ lobe bar is in grey as it derives from only a single experiment with few replicates. There were no data in the literature on the $\alpha'\beta'$ lobes or drivers that encompass all mushroom body lobes.

Figure 4. The extent of drivers' Kenyon cell expression accounts for the majority of short-term olfactory memory effects.

The estimated Kenyon cell counts for drivers were taken from Aso et al. 2009. The memory effect sizes are derived from nested, weighted, multivariate meta-regression models that adjusted for confounding variables that contributed to heterogeneity. **A.** Bubble plot of *rut* restoration; the cell count of driver lines accounts for 84% of the variance of the learning effects of rut restoration ($p < 0.0001$). Each bubble's area indicates that estimate's weight in the regression model; the blue fit line has a slope of 0.023% per cell [95CI 0.016, 0.030]. The grey line indicates the level of no rescue, i.e. the learning level of *rut* mutants. **B.** For *shi^{ts}* inactivation, 88% of the learning variance is attributable to the number of cells encompassed by the driver ($p < 0.0001$). The blue fit line has a slope of -0.034 % per cell [95CI -0.046, -0.0216]; the grey line indicates the level of no effect, i.e. the learning expected from the effect of heat alone. **C.** Learning effect per cell in mushroom body sub-regions from *rut* restoration in different lobes and combinations, adjusted for heterogeneity effects. Error bars are confidence intervals; there are no statistical differences between rut lobe categories. **D.** The *shi^{ts}* learning effect per cell in two lobes and their combination. There are no statistical differences between *shi^{ts}* lobe categories.

Figure 5. Forest plot of *rut* mutant learning changes.

Each data set is identified by the source article and figure panel. This figure is a detailed version of the same plot in the main article, but uses proportional reductions instead of percentage changes. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the percentage change relative to control.

Figure 6. Forest plot of *rut* restoration in the αβ lobes.

Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.
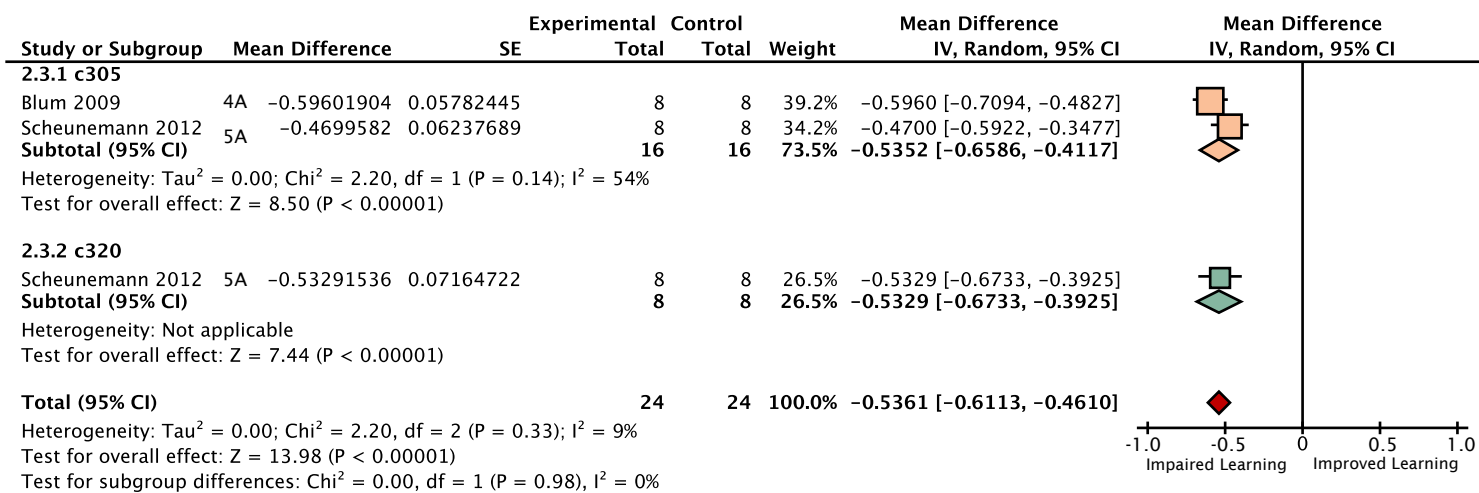
Figure 7. Forest plot of *rut* restoration in the α′β′ lobes.

Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.
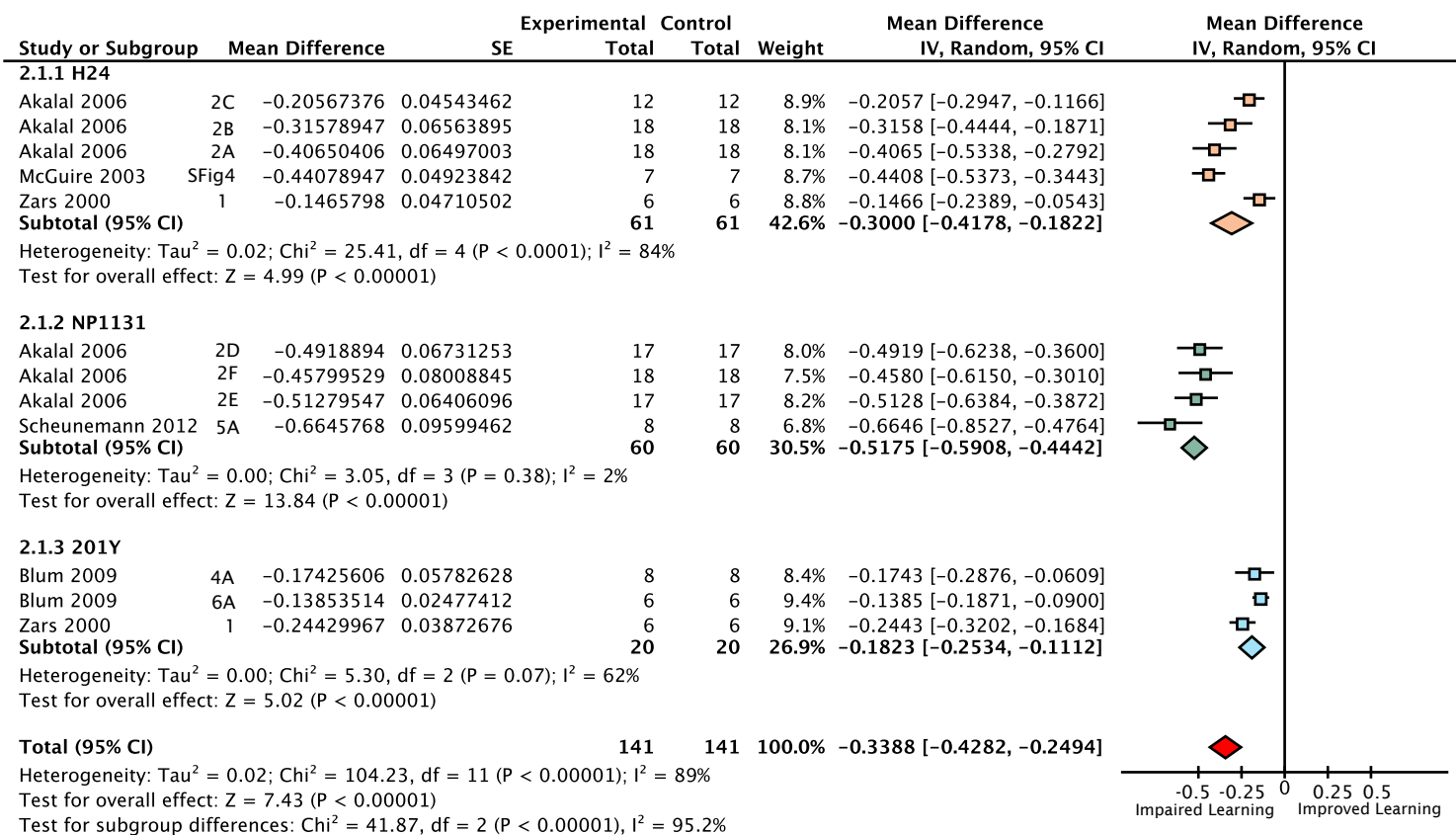
Figure 8. Forest plot of *rut* restoration in the γ lobes.

Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Figure 9. Forest plot of *rut* restoration in the αβ and γ lobes.

Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Figure 10. Forest plot of *rut* restoration in all lobes of the mushroom body.

Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Figure 11. Forest plot of the effect on STM of elevating flies from permissive to restrictive temperatures.

This figure is a detailed version of the same plot in the main article, but uses proportional reductions instead of percentage changes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Figure 12. Forest plot of experiments using *shi*[ts] to inactivate neurotransmission from the αβ lobes.

Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Figure 13. Forest plot of experiments using *shi*[ts] to inactivate neurotransmission from the αβ + γ lobes.

Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Table 1. Characteristics of included experiments.

All experiments are listed and identified by their study, figure panel and genotype/s. We name the most precise genotype possible based on the information given in the original article. Odor pair, range experimental temperature or temperature range, the nature of the conditioning shock and the relative humidity (RH) are also listed. The time delay between training and testing is listed in minutes; those labelled '0*' were reported as following training 'immediately.' Shock is listed in volts; current type is omitted if not reported in the original study. Cells containing a dash indicate that the information was not found in the original article.

Supporting Dataset 1. Meta-analyses of *rut* and *shi* STM experiments Review Manager file.

The file used to calculate and plot the meta-analyses shown in Figures 2, 3, 5-13.

Supporting Dataset 2. Spreadsheet used for meta-regression of *rut* and *shi* STM experiments.

The Excel file used as input to SAS for model construction.

Figure 1. Review overview. A. Flow chart of systematic literature review procedure. The literature was reviewed in a five stage process, starting with a PubMed search that yielded 279 articles, followed by four screens of increasing detail, reviewing the article title, abstract full text and experimental design. A total of ten articles, two of which included relevant data for both *rutabaga* and *shibire*[ts] experiments were used in the meta-analyses. **B.** Histogram of performance indices for all control experiments identified by the review.

Figure 2. Meta-analyses of *rutabaga* mutant lines and targeted transgenic restoration; short-term memory data are expressed as percentages. **A.** A summary forest plot of learning changes observed in 340 experiments with *rut* mutant lines, with subgroups showing the differences between the different *rut* alleles and strains. Learning is expressed as a percentage change relative to wild type. The red diamond on the bottom line indicates the overall impairment in learning in the *rut* hypomorphs relative to wild type controls is -60% [95CI -56%, -64%]. The complete forest plot is given in Figure 5. **B.** Summary estimates from the *rut* mutant meta-analysis and five meta-analyses of lobular restoration experiments. Learning is displayed as a percentage of wild type learning. The markers indicate the proportion of learning relative to wild type expressed as a percentage; error bars are 95% confidence intervals. To the right of the markers are numbers for the amount of rescue (R =) relative to the *rut* hypomorphs. N(E) and N(C) are the experimental and control iterations respectively. Except for the α′β′ lobes (p=0.17), all lobe categories showed a statistically significant partial rescue of learning (αβ p=0.029, γ p<1 x 10⁻⁴⁵, αβ+γ p=1.1 x 10⁻¹⁶, all lobes p<1 x 10⁻⁴⁵) when compared with *rut* learning.

A

Learning Response to an Increase in Temperature
(Percentage Change of Performance Index)

Article	Figure	N(R)	N(P)
McGuire 2001	2AB	6	6
Schwaerzel 2002	3C	6	6
Wild type (+/+)		12	12
Akalal 2006	4A	6	6
Akalal 2006	4D	10	6
Akalal 2006	4B	6	10
Akalal 2006	4E	10	10
Akalal 2006	4F	10	13
Akalal 2006	4C	9	6
McGuire 2001	2AB	6	6
McGuire 2001	2AB	5	6
McGuire 2001	2AB	6	6
Schwaerzel 2003	1E	6	6
Driver-GAL4/+		74	75
Akalal 2006	4E	10	10
Akalal 2006	4A	6	6
Akalal 2006	4F	10	13
Akalal 2006	4D	10	6
Akalal 2006	4C	9	6
Akalal 2006	4B	6	10
Dubnau 2001	3A	6	6
McGuire 2001	2AB	6	6
Schwaerzel 2002	3C	6	6
Schwaerzel 2002	3C	6	6
Schwaerzel 2003	1E	6	6
UAS-shi^ts/+		81	81
Overall Heat Effect		167	168

Impaired Learning   Improved Learning

B

Figure 3. Meta-analyses of *shibire^ts* inhibition of neurotransmission in the mushroom body lobes. Learning data are expressed as percentages. **A.** A summary forest plot of learning changes in heat treatment controls, with subgroups showing the differences between 3 types of controls. Learning is expressed as a percentage change relative to wild type. The red diamond on the bottom line indicates that the overall impairment in learning in flies exposed to elevated temperature is -17% [95CI -12%, -22%]. A complete forest plot in shown in Figure 11. **B.** Summary estimates from the heat exposure controls and three meta-analyses of lobular inactivation experiments. Colored markers correspond to diamonds in panel A. Learning at the restrictive temperature is shown as a percentage of learning at the permissive temperature; error bars are 95% confidence intervals. To the right of the markers are numbers learning impairment (Δ* =) relative to the synthetic heat effect control. N(R) and N(P) are the restrictive and permissive iterations respectively. The αβ lobes (p=0.0001) and the αβ+γ combination (p<1 x 10^-45) show statistically significant impairment while the γ lobes do not (p=0.7071). The γ lobe bar is in grey as it derives from only a single experiment with few iterations. There were no data in the literature on the α'β' lobes or drivers that encompass all mushroom body lobes.

Figure 4. The extent of drivers' Kenyon cell expression accounts for the majority of short-term olfactory memory effects. The estimated Kenyon cell counts for drivers were taken from Aso *et al.* 2009. The memory effect sizes are derived from nested, weighted, multi-variate meta-regression models that adjusted for confounding variables that contributed to heterogeneity. **A.** Bubble plot of *rut* restoration; the cell count of driver lines accounts for 84% of the variance of the learning effects of *rut* restoration (p < 0.0001). Each bubble's area indicates that estimate's weight in the regression model; the blue fit line has a slope of 0.023% per cell [95CI 0.016, 0.030]. The grey line indicates the level of no rescue, i.e. the learning level of *rut* mutants. **B.** For *shi^{ts}* inactivation, 88% of the learning variance is attributable to the number of cells encompassed by the driver (p < 0.0001). The blue fit line has a slope of -0.034 % per cell [95CI -0.046, -0.0216]; the grey line indicates the level of no effect, i.e. the learning expected from the effect of heat alone. **C.** Learning effect per cell in mushroom body sub-regions from *rut* restoration in different lobes and combinations, adjusted for heterogeneity effects. Error bars are confidence intervals; there are no statistical differences between *rut* lobe categories. **D.** The *shi^{ts}* learning effect per cell in two lobes and their combination. There are no statistical differences between *shi^{ts}* lobe categories.

Study or Subgroup	Fig	Mean Difference	SE	Experimental Total	Control Total	Weight	Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
2.6.1 rut2080								
Blum 2009	1A	−0.65972222	0.04401288	6	6	3.3%	−0.6597 [−0.7460, −0.5735]	
McGuire 2003	2A	−0.59554032	0.06932928	5	5	2.7%	−0.5955 [−0.7314, −0.4597]	
Schwaerzel 2003	1C	−0.62375664	0.04974143	6	6	3.2%	−0.6238 [−0.7212, −0.5263]	
Zars 2000	1	−0.43322476	0.07919526	6	6	2.4%	−0.4332 [−0.5884, −0.2780]	
Subtotal (95% CI)				**23**	**23**	**11.6%**	**−0.5949 [−0.6782, −0.5115]**	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 6.38, df = 3 (P = 0.09); I^2 = 53%								
Test for overall effect: Z = 13.99 (P < 0.00001)								
2.6.2 rut2080+DL								
Akalal 2006	3F	−0.64454627	0.05335415	12	12	3.1%	−0.6445 [−0.7491, −0.5400]	
Akalal 2006	2A	−0.4893617	0.06009956	12	12	2.9%	−0.4894 [−0.6072, −0.3716]	
Akalal 2006	2B	−0.63414634	0.06263107	18	18	2.8%	−0.6341 [−0.7569, −0.5114]	
Akalal 2006	5C	−0.7804878	0.09220162	6	6	2.1%	−0.7805 [−0.9612, −0.5998]	
Akalal 2006	3D	−0.75558517	0.06358167	24	24	2.8%	−0.7556 [−0.8802, −0.6310]	
Akalal 2006	3A	−0.65	0.0701181	12	12	2.7%	−0.6500 [−0.7874, −0.5126]	
Akalal 2006	5B	−0.70909091	0.04854018	6	6	3.2%	−0.7091 [−0.8042, −0.6140]	
Akalal 2006	5A	−0.73239437	0.07774699	6	6	2.5%	−0.7324 [−0.8848, −0.5800]	
Akalal 2006	2F	−0.73279246	0.07182689	18	18	2.6%	−0.7328 [−0.8736, −0.5920]	
Akalal 2006	2C	−0.58646616	0.04529384	18	18	3.3%	−0.5865 [−0.6752, −0.4977]	
Akalal 2006	3B	−0.71157238	0.01987848	10	10	3.8%	−0.7116 [−0.7505, −0.6726]	
Akalal 2006	3E	−0.71662173	0.04324871	12	12	3.3%	−0.7166 [−0.8014, −0.6319]	
Akalal 2006	3C	−0.49206349	0.09338142	12	12	2.1%	−0.4921 [−0.6751, −0.3090]	
Akalal 2006	2E	−0.7094159	0.04298548	18	18	3.3%	−0.7094 [−0.7937, −0.6252]	
Akalal 2006	2D	−0.76608295	0.0455499	17	17	3.3%	−0.7661 [−0.8554, −0.6768]	
McGuire 2003	S4	−0.57894737	0.08199867	7	7	2.4%	−0.5789 [−0.7397, −0.4182]	
McGuire 2003	S4	−0.63953488	0.05847849	7	7	3.0%	−0.6395 [−0.7542, −0.5249]	
McGuire 2003	2A	−0.6500205	0.11159574	5	5	1.8%	−0.6500 [−0.8687, −0.4313]	
McGuire 2003	2A	−0.60461582	0.07705735	5	5	2.5%	−0.6046 [−0.7556, −0.4536]	
McGuire 2003	S4	−0.7260274	0.09062861	7	7	2.2%	−0.7260 [−0.9037, −0.5484]	
Schwaerzel 2003	1C	−0.52595681	0.06428675	6	6	2.8%	−0.5260 [−0.6520, −0.4000]	
Subtotal (95% CI)				**238**	**238**	**58.5%**	**−0.6649 [−0.6997, −0.6301]**	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 39.79, df = 20 (P = 0.005); I^2 = 50%								
Test for overall effect: Z = 37.44 (P < 0.00001)								
2.6.3 rut1								
Blum 2009	1A	−0.57971014	0.06697161	6	6	2.7%	−0.5797 [−0.7110, −0.4484]	
Scheunemann 2012	5A	−0.63796611	0.05973091	8	8	2.9%	−0.6380 [−0.7550, −0.5209]	
Subtotal (95% CI)				**14**	**14**	**5.7%**	**−0.6122 [−0.6995, −0.5248]**	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 0.42, df = 1 (P = 0.52); I^2 = 0%								
Test for overall effect: Z = 13.73 (P < 0.00001)								
2.6.4 rut2080; UAS−rut								
Blum 2009	3A	−0.39694656	0.05422249	6	6	3.1%	−0.3969 [−0.5032, −0.2907]	
Blum 2009	4A	−0.51162791	0.0554394	8	8	3.0%	−0.5116 [−0.6203, −0.4030]	
Blum 2009	4B	−0.35658915	0.05628878	12	12	3.0%	−0.3566 [−0.4669, −0.2463]	
Blum 2009	6A	−0.33333333	0.08628032	6	6	2.3%	−0.3333 [−0.5024, −0.1642]	
McGuire 2003	2A	−0.59547356	0.08633171	5	5	2.3%	−0.5955 [−0.7647, −0.4263]	
Zars 2000	1	−0.37459283	0.06429222	6	6	2.8%	−0.3746 [−0.5006, −0.2486]	
Schwaerzel 2003	1C	−0.57380652	0.05505235	6	6	3.0%	−0.5738 [−0.6817, −0.4659]	
Thum 2007	1D	−0.4866342	0.12155664	8	8	1.6%	−0.4866 [−0.7249, −0.2484]	
Subtotal (95% CI)				**57**	**57**	**21.1%**	**−0.4511 [−0.5221, −0.3801]**	
Heterogeneity: Tau^2 = 0.01; Chi^2 = 16.14, df = 7 (P = 0.02); I^2 = 57%								
Test for overall effect: Z = 12.45 (P < 0.00001)								
2.6.5 rut1;UAS−rut								
Scheunemann 2012	5A	−0.50962605	0.05377106	8	8	3.1%	−0.5096 [−0.6150, −0.4042]	
Subtotal (95% CI)				**8**	**8**	**3.1%**	**−0.5096 [−0.6150, −0.4042]**	
Heterogeneity: Not applicable								
Test for overall effect: Z = 9.48 (P < 0.00001)								
Total (95% CI)				**340**	**340**	**100.0%**	**−0.6008 [−0.6400, −0.5616]**	
Heterogeneity: Tau^2 = 0.01; Chi^2 = 148.50, df = 35 (P < 0.00001); I^2 = 76%								
Test for overall effect: Z = 30.03 (P < 0.00001)								
Test for subgroup differences: Chi^2 = 32.36, df = 4 (P < 0.00001), I^2 = 87.6%								

−1.0 −0.5 0 0.5 1.0
Impaired Learning   Improved Learning

Figure 5. Forest plot of *rut* mutant learning changes. Each data set is identified by the source article and figure panel. This figure is a detailed version of the same plot in the main article, but uses proportional reductions instead of percentage changes. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the percentage change relative to control.

Study or Subgroup		Mean Difference	SE	Experimental Total	Control Total	Weight	Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
2.2.1 17d								
Akalal 2006	3D	−0.50372345	0.04501461	24	24	9.0%	−0.5037 [−0.5920, −0.4155]	
Akalal 2006	3F	−0.46681941	0.05444332	12	12	8.7%	−0.4668 [−0.5735, −0.3601]	
Akalal 2006	3E	−0.6297295	0.0837085	12	12	7.6%	−0.6297 [−0.7938, −0.4657]	
Scheunemann 2012	5A	−0.62277952	0.06617136	8	8	8.3%	−0.6228 [−0.7525, −0.4931]	
Zars 2000	1	−0.30141844	0.09124256	6	6	7.3%	−0.3014 [−0.4803, −0.1226]	
Subtotal (95% CI)				62	62	41.0%	**−0.5096 [−0.6032, −0.4160]**	

Heterogeneity: Tau^2 = 0.01; Chi^2 = 10.82, df = 4 (P = 0.03); I^2 = 63%
Test for overall effect: Z = 10.67 (P < 0.00001)

2.2.2 c739								
Akalal 2006	3B	−0.69875186	0.01996179	10	10	9.5%	−0.6988 [−0.7379, −0.6596]	
Akalal 2006	3A	−0.50714286	0.05890347	12	12	8.6%	−0.5071 [−0.6226, −0.3917]	
Akalal 2006	3C	−0.44444444	0.08944164	12	12	7.4%	−0.4444 [−0.6197, −0.2691]	
Blum 2009	4B	−0.33583882	0.03818306	12	12	9.2%	−0.3358 [−0.4107, −0.2610]	
Blum 2009	6A	−0.35598978	0.03592817	6	6	9.2%	−0.3560 [−0.4264, −0.2856]	
McGuire 2003	SFig4	−0.55232558	0.05866153	7	7	8.6%	−0.5523 [−0.6673, −0.4374]	
Subtotal (95% CI)				59	59	52.5%	**−0.4835 [−0.6429, −0.3242]**	

Heterogeneity: Tau^2 = 0.04; Chi^2 = 117.51, df = 5 (P < 0.00001); I^2 = 96%
Test for overall effect: Z = 5.95 (P < 0.00001)

2.2.3 189y								
Zars 2000	1	−0.32333333	0.1116634	6	6	6.6%	−0.3233 [−0.5422, −0.1045]	
Subtotal (95% CI)				6	6	6.6%	**−0.3233 [−0.5422, −0.1045]**	

Heterogeneity: Not applicable
Test for overall effect: Z = 2.90 (P = 0.004)

Total (95% CI)				127	127	100.0%	**−0.4830 [−0.5815, −0.3845]**	

Heterogeneity: Tau^2 = 0.03; Chi^2 = 135.06, df = 11 (P < 0.00001); I^2 = 92%
Test for overall effect: Z = 9.61 (P < 0.00001)
Test for subgroup differences: Chi^2 = 2.35, df = 2 (P = 0.31), I^2 = 15.0%

-100   -50   0   50   100
Impaired Learning   Improved Learning

Figure 6. Forest plot of *rut* restoration in the αβ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Study or Subgroup	Mean Difference		SE	Experimental Total	Control Total	Weight	Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
2.3.1 c305								
Blum 2009	4A	−0.59601904	0.05782445	8	8	39.2%	−0.5960 [−0.7094, −0.4827]	
Scheunemann 2012	5A	−0.4699582	0.06237689	8	8	34.2%	−0.4700 [−0.5922, −0.3477]	
Subtotal (95% CI)				16	16	73.5%	**−0.5352 [−0.6586, −0.4117]**	
Heterogeneity: Tau² = 0.00; Chi² = 2.20, df = 1 (P = 0.14); I² = 54%								
Test for overall effect: Z = 8.50 (P < 0.00001)								
2.3.2 c320								
Scheunemann 2012	5A	−0.53291536	0.07164722	8	8	26.5%	−0.5329 [−0.6733, −0.3925]	
Subtotal (95% CI)				8	8	26.5%	**−0.5329 [−0.6733, −0.3925]**	
Heterogeneity: Not applicable								
Test for overall effect: Z = 7.44 (P < 0.00001)								
Total (95% CI)				24	24	100.0%	**−0.5361 [−0.6113, −0.4610]**	
Heterogeneity: Tau² = 0.00; Chi² = 2.20, df = 2 (P = 0.33); I² = 9%								
Test for overall effect: Z = 13.98 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.00, df = 1 (P = 0.98), I² = 0%								

-1.0   -0.5   0   0.5   1.0
Impaired Learning   Improved Learning

Figure 7. Forest plot of *rut* restoration in the α′β′ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Study or Subgroup		Mean Difference	SE	Experimental Total	Control Total	Weight	Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
2.1.1 H24								
Akalal 2006	2C	−0.20567376	0.04543462	12	12	8.9%	−0.2057 [−0.2947, −0.1166]	
Akalal 2006	2B	−0.31578947	0.06563895	18	18	8.1%	−0.3158 [−0.4444, −0.1871]	
Akalal 2006	2A	−0.40650406	0.06497003	18	18	8.1%	−0.4065 [−0.5338, −0.2792]	
McGuire 2003	SFig4	−0.44078947	0.04923842	7	7	8.7%	−0.4408 [−0.5373, −0.3443]	
Zars 2000	1	−0.1465798	0.04710502	6	6	8.8%	−0.1466 [−0.2389, −0.0543]	
Subtotal (95% CI)				61	61	42.6%	**−0.3000 [−0.4178, −0.1822]**	
Heterogeneity: Tau^2 = 0.02; Chi^2 = 25.41, df = 4 (P < 0.0001); I^2 = 84%								
Test for overall effect: Z = 4.99 (P < 0.00001)								
2.1.2 NP1131								
Akalal 2006	2D	−0.4918894	0.06731253	17	17	8.0%	−0.4919 [−0.6238, −0.3600]	
Akalal 2006	2F	−0.45799529	0.08008845	18	18	7.5%	−0.4580 [−0.6150, −0.3010]	
Akalal 2006	2E	−0.51279547	0.06406096	17	17	8.2%	−0.5128 [−0.6384, −0.3872]	
Scheunemann 2012	5A	−0.6645768	0.09599462	8	8	6.8%	−0.6646 [−0.8527, −0.4764]	
Subtotal (95% CI)				60	60	30.5%	**−0.5175 [−0.5908, −0.4442]**	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 3.05, df = 3 (P = 0.38); I^2 = 2%								
Test for overall effect: Z = 13.84 (P < 0.00001)								
2.1.3 201Y								
Blum 2009	4A	−0.17425606	0.05782628	8	8	8.4%	−0.1743 [−0.2876, −0.0609]	
Blum 2009	6A	−0.13853514	0.02477412	6	6	9.4%	−0.1385 [−0.1871, −0.0900]	
Zars 2000	1	−0.24429967	0.03872676	6	6	9.1%	−0.2443 [−0.3202, −0.1684]	
Subtotal (95% CI)				20	20	26.9%	**−0.1823 [−0.2534, −0.1112]**	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 5.30, df = 2 (P = 0.07); I^2 = 62%								
Test for overall effect: Z = 5.02 (P < 0.00001)								
Total (95% CI)				141	141	100.0%	**−0.3388 [−0.4282, −0.2494]**	
Heterogeneity: Tau^2 = 0.02; Chi^2 = 104.23, df = 11 (P < 0.00001); I^2 = 89%								
Test for overall effect: Z = 7.43 (P < 0.00001)								
Test for subgroup differences: Chi^2 = 41.87, df = 2 (P < 0.00001), I^2 = 95.2%								

-0.5 -0.25 0 0.25 0.5
Impaired Learning   Improved Learning

Figure 8. Forest plot of *rut* restoration in the γ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Study or Subgroup		Mean Difference	SE	Experimental Total	Control Total	Weight	Mean Difference IV, Random, 95% CI
2.5.1 MB247							
Blum 2009	3A	−0.14545802	0.0374118	7	6	12.4%	−0.1455 [−0.2188, −0.0721]
McGuire 2003	2A	−0.15944085	0.09250284	5	5	2.0%	−0.1594 [−0.3407, 0.0219]
Scheunemann 2012	5A	−0.04780564	0.0944519	8	8	1.9%	−0.0478 [−0.2329, 0.1373]
Schwaerzel 2003	1C	−0.10156367	0.09239735	6	6	2.0%	−0.1016 [−0.2827, 0.0795]
Thum 2007	1D	−0.00372773	0.11317442	8	8	1.4%	−0.0037 [−0.2255, 0.2181]
Zars 2000	1	−0.01954397	0.05017942	6	6	6.9%	−0.0195 [−0.1179, 0.0788]
Subtotal (95% CI)				**40**	**39**	**26.6%**	**−0.0934 [−0.1480, −0.0388]**

Heterogeneity: Tau^2 = 0.00; Chi^2 = 5.47, df = 5 (P = 0.36); I^2 = 9%
Test for overall effect: Z = 3.35 (P = 0.0008)

2.5.2 c772							
McGuire 2003	2A	−0.08638266	0.08272642	5	5	2.5%	−0.0864 [−0.2485, 0.0758]
Zars 2000	1	−0.07166124	0.04770951	6	6	7.6%	−0.0717 [−0.1652, 0.0218]
Subtotal (95% CI)				**11**	**11**	**10.1%**	**−0.0753 [−0.1563, 0.0057]**

Heterogeneity: Tau^2 = 0.00; Chi^2 = 0.02, df = 1 (P = 0.88); I^2 = 0%
Test for overall effect: Z = 1.82 (P = 0.07)

2.5.3 c309							
Blum 2009	3A	−0.12554628	0.04941773	7	6	7.1%	−0.1255 [−0.2224, −0.0287]
Subtotal (95% CI)				**7**	**6**	**7.1%**	**−0.1255 [−0.2224, −0.0287]**

Heterogeneity: Not applicable
Test for overall effect: Z = 2.54 (P = 0.01)

2.5.4 c739+201Y							
Blum 2009	6A	−0.06663386	0.02216279	6	6	35.3%	−0.0666 [−0.1101, −0.0232]
Subtotal (95% CI)				**6**	**6**	**35.3%**	**−0.0666 [−0.1101, −0.0232]**

Heterogeneity: Not applicable
Test for overall effect: Z = 3.01 (P = 0.003)

2.5.5 c739+H24							
Akalal 2006	5A	0.05633803	0.07107945	6	6	3.4%	0.0563 [−0.0830, 0.1957]
Akalal 2006	5C	−0.09756098	0.03499166	6	6	14.2%	−0.0976 [−0.1661, −0.0290]
Akalal 2006	5B	0.03636364	0.13470832	6	6	1.0%	0.0364 [−0.2277, 0.3004]
McGuire 2003	SFig4	−0.13013699	0.08698619	7	7	2.3%	−0.1301 [−0.3006, 0.0404]
Subtotal (95% CI)				**25**	**25**	**20.8%**	**−0.0527 [−0.1412, 0.0358]**

Heterogeneity: Tau^2 = 0.00; Chi^2 = 4.88, df = 3 (P = 0.18); I^2 = 39%
Test for overall effect: Z = 1.17 (P = 0.24)

Total (95% CI)				**89**	**87**	**100.0%**	**−0.0802 [−0.1060, −0.0544]**

Heterogeneity: Tau^2 = 0.00; Chi^2 = 12.13, df = 13 (P = 0.52); I^2 = 0%
Test for overall effect: Z = 6.09 (P < 0.00001)
Test for subgroup differences: Chi^2 = 1.81, df = 4 (P = 0.77), I^2 = 0%

Figure 9. Forest plot of *rut* restoration in the αβ and γ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Study or Subgroup	Mean Difference		SE	Experimental Total	Control Total	Weight	Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
2.4.1 OK107								
Blum 2009	3A	−0.00149332	0.0412544	7	6	32.0%	−0.0015 [−0.0824, 0.0794]	
Scheunemann 2012	5A	−0.09874608	0.09392339	8	8	9.0%	−0.0987 [−0.2828, 0.0853]	
Subtotal (95% CI)				15	14	40.9%	−0.0172 [−0.0913, 0.0568]	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 0.90, df = 1 (P = 0.34); I^2 = 0%								
Test for overall effect: Z = 0.46 (P = 0.65)								
2.4.2 238y								
Zars 2000	1	−0.11074919	0.0432971	6	6	30.1%	−0.1107 [−0.1956, −0.0259]	
Subtotal (95% CI)				6	6	30.1%	−0.1107 [−0.1956, −0.0259]	
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.56 (P = 0.01)								
2.4.3 30y								
Zars 2000	1	−0.10423453	0.04458377	6	6	29.0%	−0.1042 [−0.1916, −0.0169]	
Subtotal (95% CI)				6	6	29.0%	−0.1042 [−0.1916, −0.0169]	
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.34 (P = 0.02)								
Total (95% CI)				27	26	100.0%	−0.0729 [−0.1313, −0.0144]	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 4.32, df = 3 (P = 0.23); I^2 = 31%								
Test for overall effect: Z = 2.44 (P = 0.01)								
Test for subgroup differences: Chi^2 = 3.43, df = 2 (P = 0.18), I^2 = 41.6%								

-0.2 -0.1 0 0.1 0.2
Impaired Learning   Improved Learning

Figure 10. Forest plot of *rut* restoration in all lobes of the mushroom body. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Study or Subgroup		Mean Difference	SE	Change in learning Total	Control Total	Weight	Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
4.1.1 Wild type								
McGuire 2001	2AB	-0.098756	0.065031	6	6	5.0%	-0.0988 [-0.2262, 0.0287]	
Schwaerzel 2002	3C	-0.302375	0.074808	6	6	4.6%	-0.3024 [-0.4490, -0.1558]	
Subtotal (95% CI)				12	12	9.6%	-0.1972 [-0.3966, 0.0022]	
Heterogeneity: Tau^2 = 0.02; Chi^2 = 4.22, df = 1 (P = 0.04); I^2 = 76%								
Test for overall effect: Z = 1.94 (P = 0.05)								
4.1.2 Driver line								
Akalal 2006	4A	-0.157895	0.098759	6	6	3.6%	-0.1579 [-0.3515, 0.0357]	
Akalal 2006	4D	-0.198862	0.075507	10	6	4.5%	-0.1989 [-0.3469, -0.0509]	
Akalal 2006	4B	-0.215819	0.088508	6	10	4.0%	-0.2158 [-0.3893, -0.0423]	
Akalal 2006	4E	-0.198276	0.095272	10	10	3.7%	-0.1983 [-0.3850, -0.0115]	
Akalal 2006	4F	0.00877	0.0905	10	13	3.9%	0.0088 [-0.1686, 0.1861]	
Akalal 2006	4C	-0.065043	0.054632	9	6	5.5%	-0.0650 [-0.1721, 0.0420]	
McGuire 2001	2AB	-0.109932	0.057073	5	6	5.4%	-0.1099 [-0.2218, 0.0019]	
McGuire 2001	2AB	-0.162059	0.118246	6	6	2.9%	-0.1621 [-0.3938, 0.0697]	
McGuire 2001	2AB	-0.075809	0.054442	6	6	5.5%	-0.0758 [-0.1825, 0.0309]	
Schwaerzel 2003	1E	-0.184743	0.103029	6	6	3.4%	-0.1847 [-0.3867, 0.0172]	
Subtotal (95% CI)				74	75	42.5%	-0.1175 [-0.1639, -0.0711]	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 7.32, df = 9 (P = 0.60); I^2 = 0%								
Test for overall effect: Z = 4.96 (P < 0.00001)								
4.1.3 UAS Shi parental								
Akalal 2006	4F	-0.125	0.064096	10	13	5.1%	-0.1250 [-0.2506, 0.0006]	
Akalal 2006	4E	0.00917	0.104966	10	10	3.3%	0.0092 [-0.1966, 0.2149]	
Akalal 2006	4C	-0.263969	0.102135	9	6	3.4%	-0.2640 [-0.4641, -0.0638]	
Akalal 2006	4A	-0.258056	0.063261	6	6	5.1%	-0.2581 [-0.3820, -0.1341]	
Akalal 2006	4D	-0.249962	0.063705	10	6	5.1%	-0.2500 [-0.3748, -0.1251]	
Akalal 2006	4B	-0.321155	0.064892	6	10	5.0%	-0.3212 [-0.4483, -0.1940]	
Dubnau 2001	3A	0.123	0.141187	6	6	2.3%	0.1230 [-0.1537, 0.3997]	
McGuire 2001	2AB	-0.185853	0.065479	6	6	5.0%	-0.1859 [-0.3142, -0.0575]	
Schwaerzel 2002	3C	-0.376934	0.047564	6	6	5.9%	-0.3769 [-0.4702, -0.2837]	
Schwaerzel 2002	3C	-0.29462	0.072966	6	6	4.6%	-0.2946 [-0.4376, -0.1516]	
Schwaerzel 2003	1E	0.09119	0.116895	6	6	3.0%	0.0912 [-0.1379, 0.3203]	
Subtotal (95% CI)				81	81	47.9%	-0.1968 [-0.2794, -0.1142]	
Heterogeneity: Tau^2 = 0.01; Chi^2 = 34.73, df = 10 (P = 0.0001); I^2 = 71%								
Test for overall effect: Z = 4.67 (P < 0.00001)								
Total (95% CI)				167	168	100.0%	-0.1706 [-0.2217, -0.1195]	
Heterogeneity: Tau^2 = 0.01; Chi^2 = 60.21, df = 22 (P < 0.0001); I^2 = 63%								
Test for overall effect: Z = 6.54 (P < 0.00001)								
Test for subgroup differences: Chi^2 = 3.03, df = 2 (P = 0.22), I^2 = 34.0%								

-0.5   -0.25   0   0.25   0.50
Impaired Learning   Improved Learning

Figure 11. Forest plot of the effect on STM from elevating flies from permissive to restrictive temperatures. This figure is a detailed version of the same plot in the main article, but uses proportional reductions instead of percentage changes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.
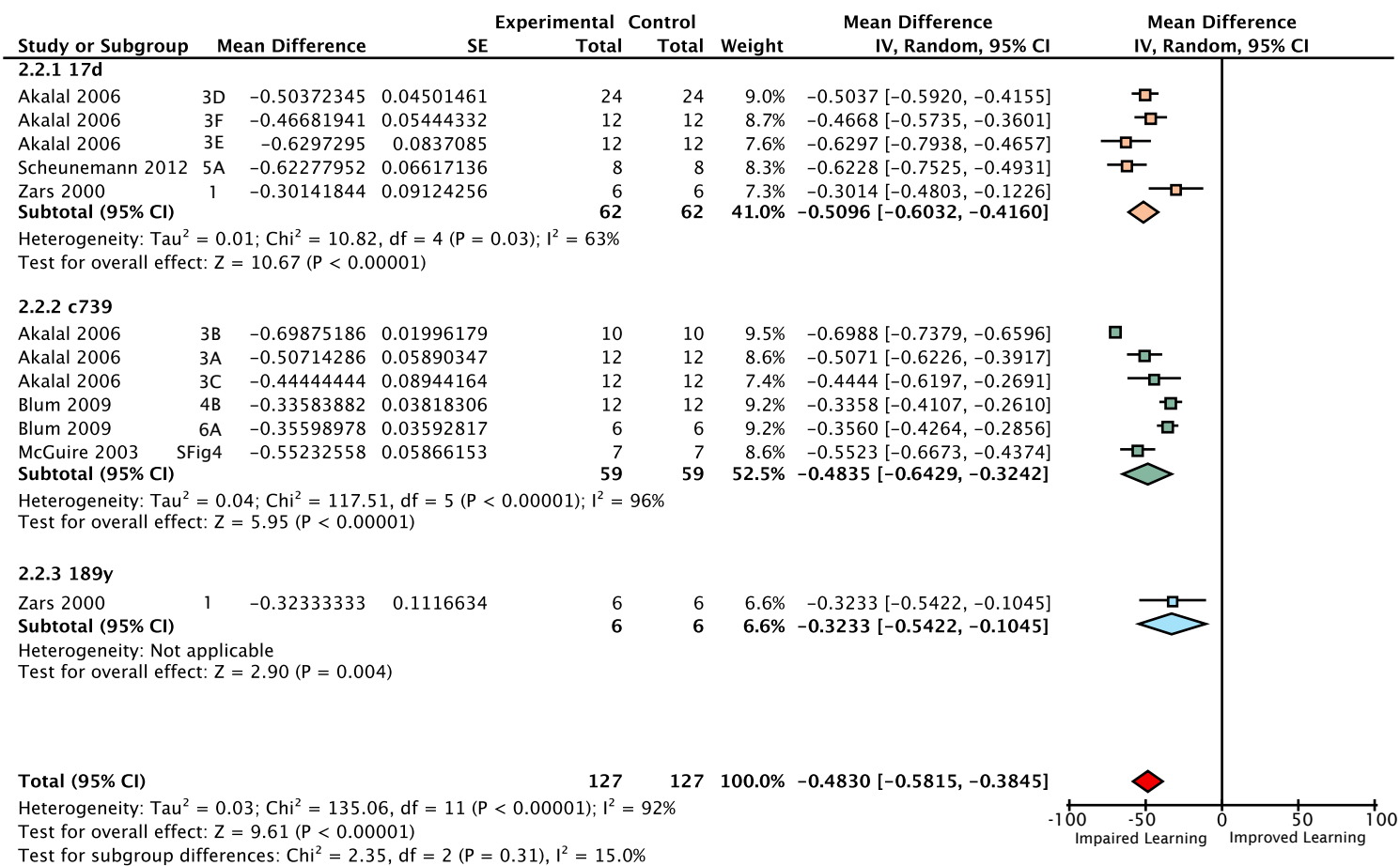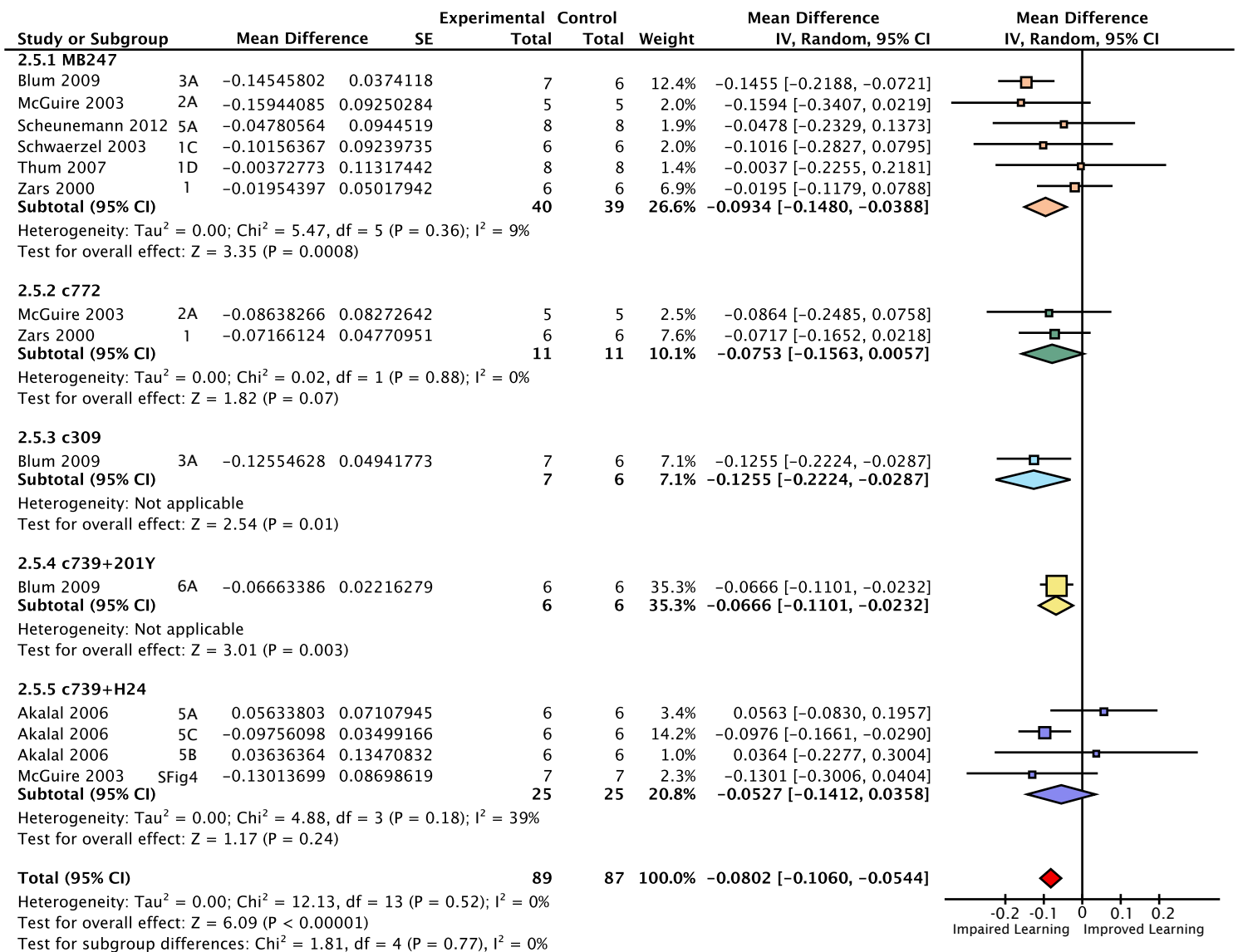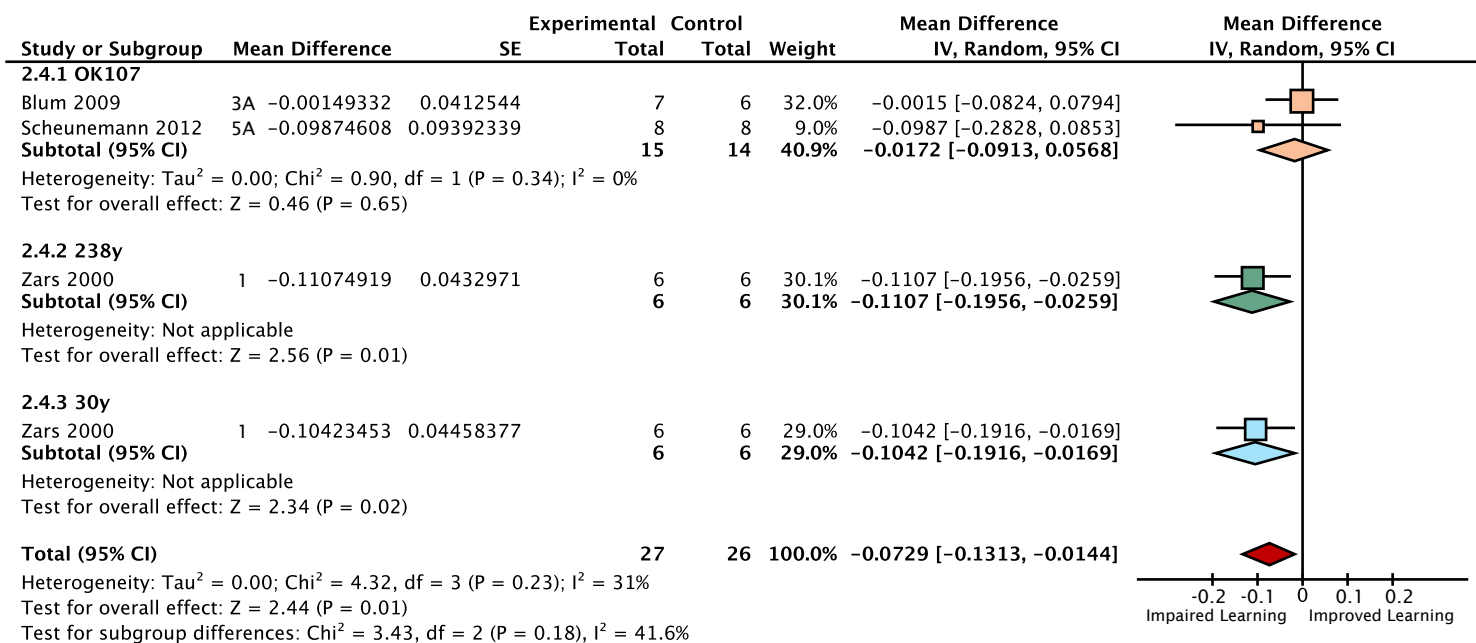
Study or Subgroup		Mean Difference	SE	Experimental Total	Control Total	Weight	Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
6.2.1 17d								
Akalal 2006	4E	−0.464236	0.047668	10	10	15.3%	−0.46 [−0.56, −0.37]	
Akalal 2006	4F	−0.054545	0.06666	13	10	14.0%	−0.05 [−0.19, 0.08]	
Akalal 2006	4D	−0.314286	0.120816	6	10	10.1%	−0.31 [−0.55, −0.08]	
Subtotal (95% CI)				**29**	**30**	**39.5%**	**−0.28 [−0.57, 0.01]**	

Heterogeneity: Tau^2 = 0.06; Chi^2 = 25.00, df = 2 (P < 0.00001); I^2 = 92%
Test for overall effect: Z = 1.89 (P = 0.06)

6.2.2 c739								
Akalal 2006	4C	−0.554766	0.033392	6	9	16.1%	−0.55 [−0.62, −0.49]	
Akalal 2006	4A	−0.476173	0.063252	6	6	14.3%	−0.48 [−0.60, −0.35]	
Akalal 2006	4B	−0.58275	0.048244	10	6	15.3%	−0.58 [−0.68, −0.49]	
McGuire 2001	2AB	−0.456296	0.054318	6	6	14.9%	−0.46 [−0.56, −0.35]	
Subtotal (95% CI)				**28**	**27**	**60.5%**	**−0.53 [−0.58, −0.47]**	

Heterogeneity: Tau^2 = 0.00; Chi^2 = 4.29, df = 3 (P = 0.23); I^2 = 30%
Test for overall effect: Z = 18.55 (P < 0.00001)

Total (95% CI)				**57**	**57**	**100.0%**	**−0.42 [−0.54, −0.31]**	

Heterogeneity: Tau^2 = 0.02; Chi^2 = 52.42, df = 6 (P < 0.00001); I^2 = 89%
Test for overall effect: Z = 6.98 (P < 0.00001)
Test for subgroup differences: Chi^2 = 2.79, df = 1 (P = 0.10), I^2 = 64.1%

Figure 12. Forest plot of experiments using *shi*^{ts} to inactivate neurotransmission from the αβ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.
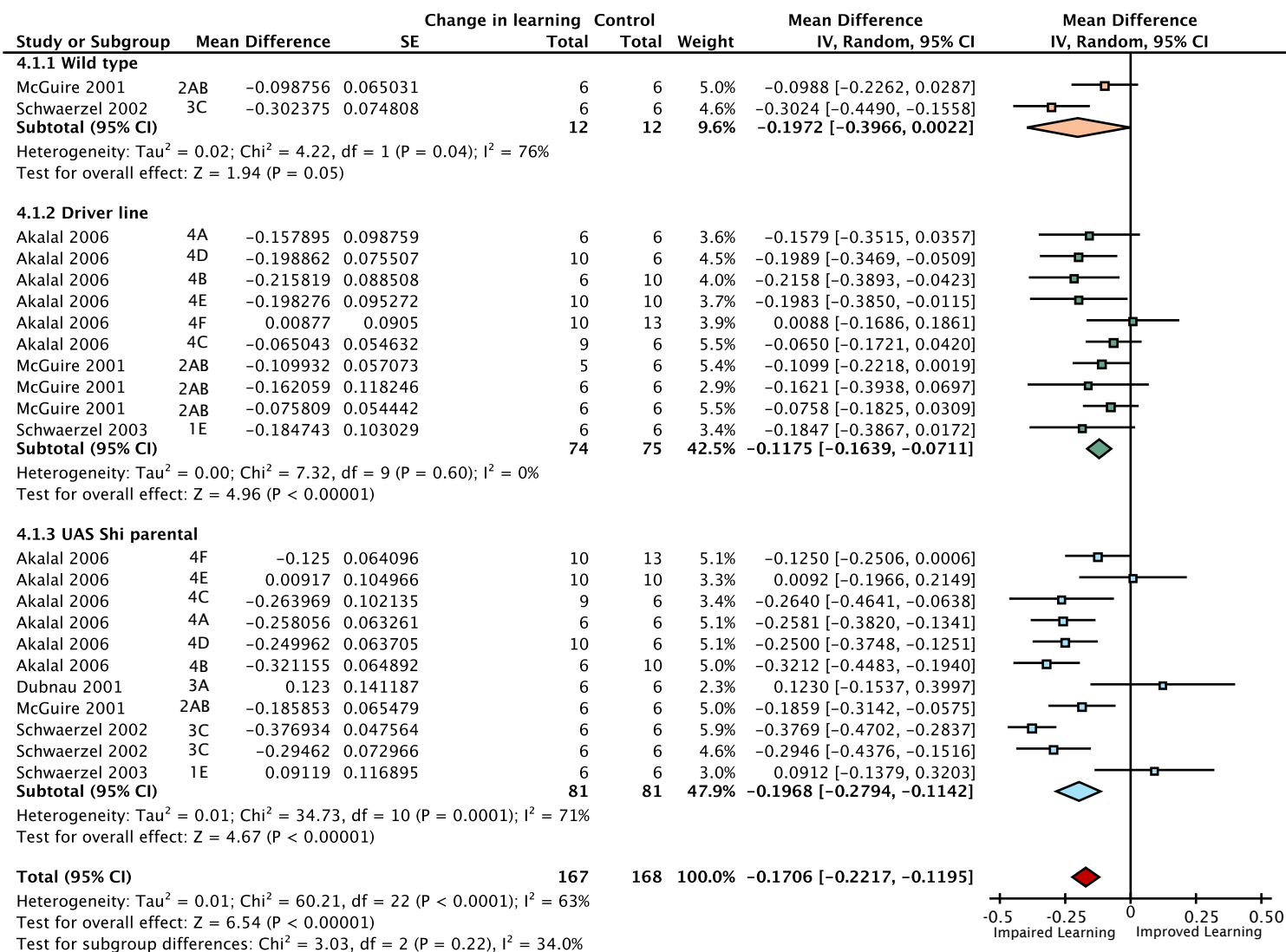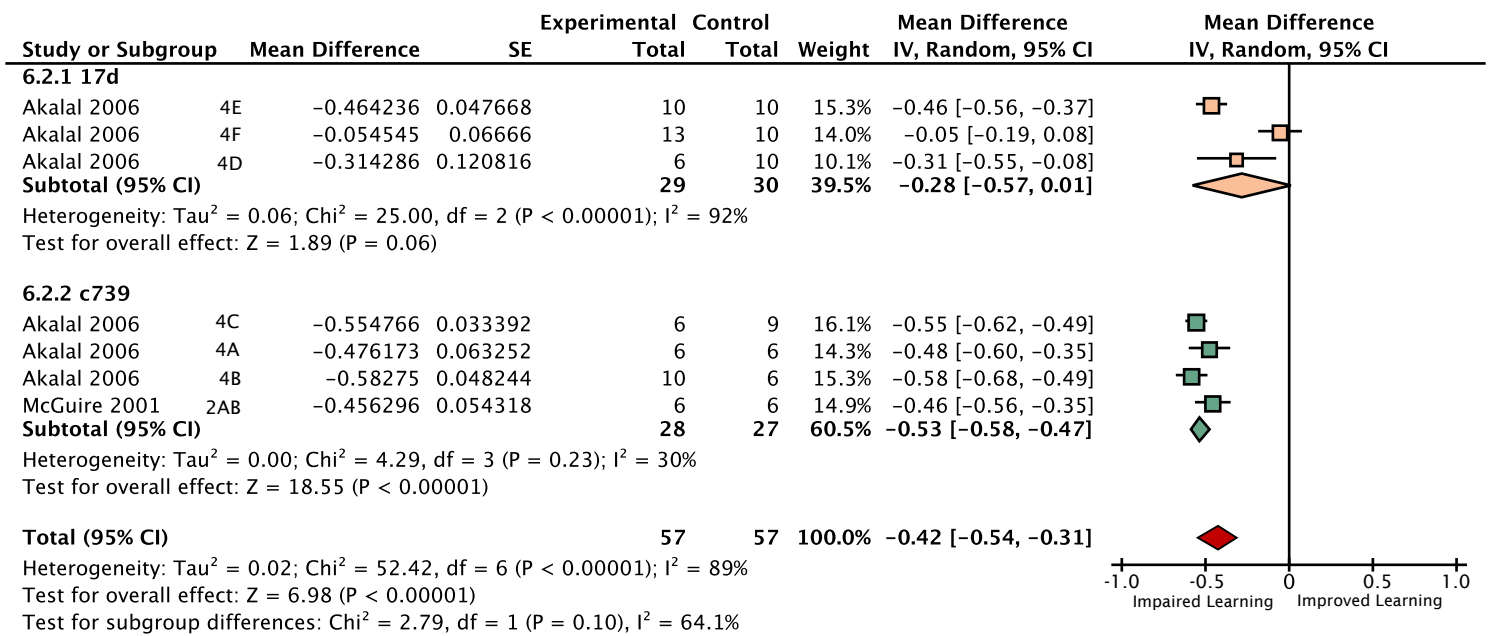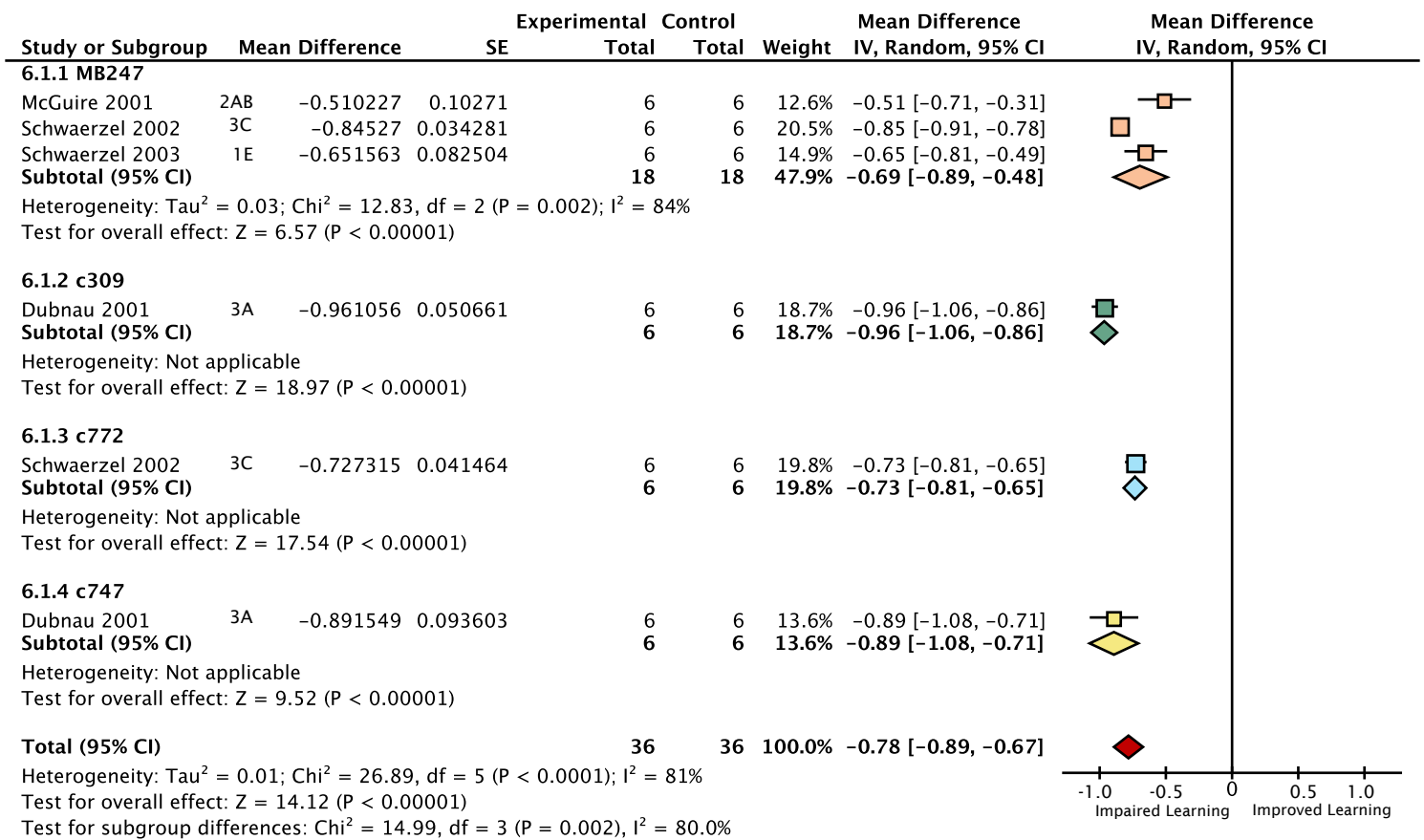
Study or Subgroup		Mean Difference	SE	Experimental Total	Control Total	Weight	Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
6.1.1 MB247								
McGuire 2001	2AB	−0.510227	0.10271	6	6	12.6%	−0.51 [−0.71, −0.31]	
Schwaerzel 2002	3C	−0.84527	0.034281	6	6	20.5%	−0.85 [−0.91, −0.78]	
Schwaerzel 2003	1E	−0.651563	0.082504	6	6	14.9%	−0.65 [−0.81, −0.49]	
Subtotal (95% CI)				**18**	**18**	**47.9%**	**−0.69 [−0.89, −0.48]**	
Heterogeneity: Tau² = 0.03; Chi² = 12.83, df = 2 (P = 0.002); I² = 84%								
Test for overall effect: Z = 6.57 (P < 0.00001)								
6.1.2 c309								
Dubnau 2001	3A	−0.961056	0.050661	6	6	18.7%	−0.96 [−1.06, −0.86]	
Subtotal (95% CI)				**6**	**6**	**18.7%**	**−0.96 [−1.06, −0.86]**	
Heterogeneity: Not applicable								
Test for overall effect: Z = 18.97 (P < 0.00001)								
6.1.3 c772								
Schwaerzel 2002	3C	−0.727315	0.041464	6	6	19.8%	−0.73 [−0.81, −0.65]	
Subtotal (95% CI)				**6**	**6**	**19.8%**	**−0.73 [−0.81, −0.65]**	
Heterogeneity: Not applicable								
Test for overall effect: Z = 17.54 (P < 0.00001)								
6.1.4 c747								
Dubnau 2001	3A	−0.891549	0.093603	6	6	13.6%	−0.89 [−1.08, −0.71]	
Subtotal (95% CI)				**6**	**6**	**13.6%**	**−0.89 [−1.08, −0.71]**	
Heterogeneity: Not applicable								
Test for overall effect: Z = 9.52 (P < 0.00001)								
Total (95% CI)				**36**	**36**	**100.0%**	**−0.78 [−0.89, −0.67]**	
Heterogeneity: Tau² = 0.01; Chi² = 26.89, df = 5 (P < 0.0001); I² = 81%								
Test for overall effect: Z = 14.12 (P < 0.00001)								
Test for subgroup differences: Chi² = 14.99, df = 3 (P = 0.002), I² = 80.0%								

Figure 13. Forest plot of experiments using *shi^ts* to inactivate neurotransmission from the αβ + γ lobes. Each data set is identified by the source article and figure panel. The subgroups are different driver lines, the red diamond indicates the overall estimated value range for the proportional change relative to control.

Table 1. Characteristics of included experiments. All experiments are listed and identified by their study, figure panel and genotype/s. We name the most precise genotype possible based on the information given in the original article. Odor pair, range experimental temperature or temperature range, the nature of the conditioning shock and the relative humidity (RH) are also listed. The time delay between training and testing is listed in minutes; those labelled '0*' were reported as following training 'immediately.' Shock is listed in volts, current type is omitted if not reported in the original study. Cells containing a dash indicate that the information was not found in the original article.

Study	Fig.	Genotype, Experimental	Genotype, Control	N(E)	N(C)	Odor Pair	Experimental Temp. °C	Shock (V)	Time (min)	RH (%)
rutabaga rescue in the αβ (alphabeta) lobes										
Zars 2000	1	*rut²⁰⁸⁰/Y; 17d; UAS-rut*	*17d/+*	6	6	MCH-BEN	25	120 AC	2	-
Zars 2000	1	*rut²⁰⁸⁰/Y; 189Y; UAS-rut*	*189Y/+*	6	6	MCH-BEN	25	120 AC	2	-
McGuire 2003	S4	*rut²⁰⁸⁰; c739; UAS-rut*	*c739/+*	7	7	OCT-BEN	25	90	3	-
Akalal 2006	3A	*rut²⁰⁸⁰; c739; UAS-rut*	*c739/+*	12	12	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	3B	*rut²⁰⁸⁰; c739; UAS-rut*	*c739/+*	10	10	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	3C	*rut²⁰⁸⁰; c739; UAS-rut*	*c739/+*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	3D	*rut²⁰⁸⁰; 17d; UAS-rut*	*17d/+*	24	24	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	3E	*rut²⁰⁸⁰; 17d; UAS-rut*	*17d/+*	12	12	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	3F	*rut²⁰⁸⁰; 17d; UAS-rut*	*17d/+*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Blum 2009	4B	*rut²⁰⁸⁰/Y; c739; UAS-rut*	*+/rut²⁰⁸⁰;+;UAS-rut*	12	12	MCH-OCT	22	60	2	50
Blum 2009	6A	*rut²⁰⁸⁰/Y; c739; UAS-rut*	*+/rut²⁰⁸⁰;+;UAS-rut*	6	6	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut¹; 17d; UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
rutabaga rescue in the α'β' (prime) lobes										
Blum 2009	4A	*rut²⁰⁸⁰/Y; c305a; UAS-rut*	*+/rut²⁰⁸⁰;+;UAS-rut*	8	8	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut¹; c305a; UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
Scheunemann 2012	5A	*rut¹; c320; UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
rutabaga rescue in the γ (gamma) lobes										
Zars 2000	1	*rut²⁰⁸⁰/Y; +; H24/UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
Zars 2000	1	*rut²⁰⁸⁰/Y; 201Y; UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
McGuire 2003	S4	*rut²⁰⁸⁰/Y; +; H24/UAS-rut*	*+; H24*	7	7	OCT-BEN	25	90	3	-
Akalal 2006	2A	*rut²⁰⁸⁰/Y; +; H24/UAS-rut*	*+; H24*	18	18	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	2B	*rut²⁰⁸⁰/Y; +; H24/UAS-rut*	*+; H24*	18	18	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	2C	*rut²⁰⁸⁰/Y; +; H24/UAS-rut*	*+; H24*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	2D	*rut²⁰⁸⁰; NP1131; UAS-rut*	*+; NP1131*	17	17	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	2E	*rut²⁰⁸⁰; NP1131; UAS-rut*	*+; NP1131*	17	17	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	2F	*rut²⁰⁸⁰; NP1131; UAS-rut*	*+; NP1131*	18	18	OCT-BEN	21–25	90 DC	3	60–68
Blum 2009	4A	*rut²⁰⁸⁰/Y; 201Y; UAS-rut*	*+/rut²⁰⁸⁰; +; UAS-rut*	8	8	MCH-OCT	22	60	2	50
Blum 2009	6A	*rut²⁰⁸⁰/Y; 201Y; UAS-rut*	*+/rut²⁰⁸⁰; +; UAS-rut*	6	6	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut¹; NP1131; UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
rutabaga rescue in all lobes										
Zars 2000	1	*rut²⁰⁸⁰/Y; 30Y/UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
Zars 2000	1	*rut²⁰⁸⁰/Y; 238Y/UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
Blum 2009	3A	*rut²⁰⁸⁰; +; UAS-rut; OK107*	*+/rut²⁰⁸⁰;+;UAS-rut*	7	6	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut¹; +; UAS-rut; OK107*	wild type	8	8	EA-IA	24	120 AC	3	70
rutabaga rescue in the αβγ lobes										
Zars 2000	1	*rut²⁰⁸⁰/Y; +; MB247/UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
Zars 2000	1	*rut²⁰⁸⁰/Y; c772; UAS-rut*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
McGuire 2003	2A	*rut²⁰⁸⁰; +; MB247/UAS-rut*	Canton-S	5	5	OCT-BEN	25	90	3	-
McGuire 2003	2A	*rut²⁰⁸⁰; c772; UAS-rut*	Canton-S	5	5	OCT-BEN	25	90	3	-
McGuire 2003	S4	*rut²⁰⁸⁰; c739; H24/UAS-rut*	*+; c739; H24*	7	7	OCT-BEN	25	90	3	-
Schwaerzel 2003	1C	*rut²⁰⁸⁰; UAS-rut; MB247*	Canton-S	6	6	EA-IA	26	130	3	80
Akalal 2006	5A	*rut²⁰⁸⁰; c739; H24/UAS-rut*	*+; c739; H24*	6	6	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	5B	*rut²⁰⁸⁰; c739; H24/UAS-rut*	*+; c739; H24*	6	6	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	5C	*rut²⁰⁸⁰; c739; H24/UAS-rut*	*+; c739; H24*	6	6	OCT-BEN	21–25	90 DC	3	60–68
Thum 2007	1D	*rut²⁰⁸⁰; +; MB247/UAS-rut*	*MB247/+*	8	8	MCH-OCT	25	90 DC	0*	-
Blum 2009	3A	*rut²⁰⁸⁰/Y; +; MB247/UAS-rut*	*+/rut²⁰⁸⁰; +; UAS-rut*	7	6	MCH-OCT	22	60	2	50
Blum 2009	3A	*rut²⁰⁸⁰/Y; c309; UAS-rut*	*+/rut²⁰⁸⁰; +; UAS-rut*	7	6	MCH-OCT	22	60	2	50
Blum 2009	6A	*rut²⁰⁸⁰/Y; c739/201Y; UAS-rut*	*+/rut²⁰⁸⁰; +; UAS-rut*	6	6	MCH-OCT	22	60	2	50
Scheunemann 2012	5A	*rut¹; MB247/UAS-rut*	wild type	8	8	EA-IA	24	120 AC	3	70
rutabaga²⁰⁸⁰										
Zars 2000	1	*rut2080*	Canton-S	6	6	MCH-BEN	25	120 AC	2	-
McGuire 2003	2A	*rut2080*	Canton-S	5	5	OCT-BEN	25	90	3	-
Schwaerzel 2003	1C	*rut2080*	Canton-S	6	6	EA-IA	26	130	3	80
Blum 2009	1A	*rut2080*	*rut²⁰⁸⁰/ +*	6	6	MCH-OCT	22	60	2	50
rutabaga²⁰⁸⁰ with *Driver/s*										
Akalal 2006	3C	*rut²⁰⁸⁰; c739*	*c739/+*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	2C	*rut²⁰⁸⁰; H24*	*+; H24*	18	18	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	3F	*rut²⁰⁸⁰; 17d*	*17d/+*	12	12	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	5C	*rut²⁰⁸⁰; c739; H24*	*+; c739; H24*	6	6	OCT-BEN	21–25	90 DC	3	60–68
Akalal 2006	2A	*rut²⁰⁸⁰; H24*	*+; H24*	12	12	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	2E	*rut²⁰⁸⁰; NP1131*	*+; NP1131*	18	18	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	3A	*rut²⁰⁸⁰; c739*	*c739/+*	12	12	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	5A	*rut²⁰⁸⁰; c739; H24*	*+; c739; H24*	6	6	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	3B	*rut²⁰⁸⁰; c739*	*c739/+*	10	10	MCH-OCT	21–25	90 DC	3	60–68
Akalal 2006	2D	*rut²⁰⁸⁰; NP1131*	*+; NP1131*	17	17	MCH-BEN	21–25	90 DC	3	60–68
Akalal 2006	3E	*rut²⁰⁸⁰; 17d*	*17d/+*	12	12	MCH-OCT	21–25	90 DC	3	60–68

Study	Fig	Genotype	Control	n	n	Odor	Rest.	Perm.	V	N	Range
Akalal 2006	3D	rut^2080; 17d	17d/+	24	24	MCH-BEN		21–25	90 DC	3	60–68
Akalal 2006	2F	rut^2080; NP1131	+; NP1131	18	18	OCT-BEN		21–25	90 DC	3	60–68
Akalal 2006	5B	rut^2080; c739; H24	+; c739; H24	6	6	MCH-OCT		21–25	90 DC	3	60–68
Akalal 2006	2B	rut^2080; H24	+; H24	18	18	MCH-OCT		21–25	90 DC	3	60–68
McGuire 2003	2A	rut^2080; c772	Canton-S	5	5	OCT-BEN		25	90	3	-
McGuire 2003	2A	rut^2080; MB247	Canton-S	5	5	OCT-BEN		25	90	3	-
McGuire 2003	S4	rut^2080; c379	c739/+	7	7	OCT-BEN		25	90	3	-
McGuire 2003	S4	rut^2080; H24	+; H24	7	7	OCT-BEN		25	90	3	-
McGuire 2003	S4	rut^2080; c379; H24	+; c739; H24	7	7	OCT-BEN		25	90	3	-
Schwaerzel 2003	1C	rut^2080; MB247	Canton-S	6	6	EA-IA		26	130	3	80
rutabaga^2080; UAS-rut											
Blum 2009	3A	rut^2080; +; UAS-rut	+/rut^2080; +; UAS-rut	6	6	MCH-OCT		22	60	2	50
Blum 2009	4A	rut^2080; +; UAS-rut	+/rut^2080; +; UAS-rut	8	8	MCH-OCT		22	60	2	50
Blum 2009	4B	rut^2080; +; UAS-rut	+/rut^2080; +; UAS-rut	12	12	MCH-OCT		22	60	2	50
Blum 2009	6A	rut^2080; +; UAS-rut	+/rut^2080; +; UAS-rut	6	6	MCH-OCT		22	60	2	50
McGuire 2003	2A	rut^2080; +; UAS-rut	Canton-S	5	5	OCT-BEN		25	90	3	-
Schwaerzel 2003	1C	rut^2080; +; UAS-rut	Canton-S	6	6	EA-IA		26	130	3	80
Thum 2007	1D	rut^2080; +; UAS-rut	MB247/+	6	6	MCH-OCT		25	90 DC	0*	-
Zars 2000	1	rut^2080; +; UAS-rut	Canton-S	8	8	MCH-BEN		25	120 AC	2	-
rutabaga^1; UAS-rut											
Scheunemann 2012	5A	rut^1; +; UAS-rut	wild type	8	8	EA-IA		24	120 AC	3	70
rutabaga^1											
Blum 2009	1A	rut^1	rut^1/+	6	6	MCH-OCT		22	60	2	50
Scheunemann 2012	5A	rut^1	wild type	8	8	EA-IA		24	120 AC	3	70
UAS-shibire^ts inactivation of the αβ (alphabeta) lobes							**Rest.**	**Perm.**			
McGuire 2001	2AB	c739; UAS-shi^ts1		6	6	OCT-BEN	32	25	90	3	-
Akalal 2006	4A	c739; UAS-shi^ts1		6	6	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4B	c739; UAS-shi^ts1		10	6	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4C	c739; UAS-shi^ts1		6	9	OCT-BEN	32–35	21–25	90DC	3	60–68
Akalal 2006	4D	17d; UAS-shi^ts1		6	10	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4E	17d; UAS-shi^ts1		10	10	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4F	17d; UAS-shi^ts1		13	10	OCT-BEN	32–35	21–25	90 DC	3	60–68
UAS-shibire^ts inactivation of the αβγ lobes											
Dubnau 2001	3A	UAS-shi^ts1/c309		6	6	MCH-OCT	30	20	-	0*	-
Dubnau 2001	3A	UAS-shi^ts1/c747		6	6	MCH-OCT	30	20	-	0*	-
McGuire 2001	2AB	MB247; UAS-shi^ts1		6	6	OCT-BEN	32	25	90	3	-
Schwaerzel 2002	3C	MB247/UAS-shi^ts1		6	6	OCT-BEN	34	26	-	3	85
Schwaerzel 2002	3C	c772/UAS-shi^ts2		6	6	OCT-BEN	34	26	-	3	85
Schwaerzel 2003	1E	MB247/UAS-shi^ts1		6	6	EA-IA	34	26	130	3	80
UAS-shibire^ts inactivation of the γ (gamma) lobes											
McGuire 2001	2AB	201Y; UAS-shi^ts1		3	4	OCT-BEN	32	25	90	3	-
Wild type heat effect controls											
McGuire 2001	2AB	wCS10		6	6	OCT-BEN	32	25	90	3	-
Schwaerzel 2002	3C	Canton-S		6	6	OCT-BEN	34	26	-	3	85
Driver heat effect controls											
Akalal 2006	4A	17d/+		6	6	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4D	17d/+		10	6	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4B	c739/+		6	10	OCT-BEN	32–35	21–25	90DC	3	60–68
Akalal 2006	4E	17d/+		10	10	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4F	c739/+		10	13	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4C	c739/+		9	6	OCT-BEN	32–35	21–25	90 DC	3	60–68
McGuire 2001	2AB	201Y		5	6	OCT-BEN	32	25	90	3	-
McGuire 2001	2AB	c739		6	6	OCT-BEN	32	25	90	3	-
McGuire 2001	2AB	247		6	6	OCT-BEN	32	25	90	3	-
Schwaerzel 2003	1E	247/+		6	6	EA-IA	34	26	130	3	80
UAS-shi^ts heat effect controls											
Akalal 2006	4E	w; UAS-shi^ts1		10	10	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4A	w; UAS-shi^ts1		6	6	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4F	w; UAS-shi^ts1		10	13	OCT-BEN	32–35	21–25	90DC	3	60–68
Akalal 2006	4D	w; UAS-shi^ts1		10	6	MCH-BEN	32–35	21–25	90 DC	3	60–68
Akalal 2006	4C	w; UAS-shi^ts1		9	6	MCH-OCT	32–35	21–25	90 DC	3	60–68
Akalal 2006	4B	w; UAS-shi^ts1		6	10	OCT-BEN	32–35	21–25	90 DC	3	60–68
Dubnau 2001	3A	shi^ts1/+		6	6	MCH-OCT	30	20	-	0*	-
McGuire 2001	2AB	UAS-shi^ts1		6	6	OCT-BEN	32	25	90	3	-
Schwaerzel 2002	3C	UAS-shi^ts2/+		6	6	OCT-BEN	34	26	-	3	85
Schwaerzel 2002	3C	UAS-shi^ts1/+		6	6	OCT-BEN	34	26	-	3	85
Schwaerzel 2003	1E	UAS-shi^ts1/+		6	6	EA-IA	34	26	130	3	80