Behavioral/Cognitive

Task-Dependent Behavioral Dynamics Make the Case for Temporal Integration in Multiple Strategies during Odor Processing

Donald E. Frederick,[1,2] Austin Brown,[2,3] Stephanie Tacopina,[2] Nisarg Mehta,[2] Mark Vujovic,[2] Elizabeth Brim,[3] Tasneem Amina,[3] Bethany Fixsen,[3] and Leslie M. Kay[1,2,3]

[1]Department of Psychology, [2]Institute for Mind and Biology, and [3]The College, University of Chicago, Chicago, Illinois 60637

Differing results in olfactory-based decision-making research regarding the amount of time that rats and mice use to identify odors have led to some disagreements about odor-processing mechanics, including whether or not rodents use temporal integration (i.e., sniffing longer to identify odors better). Reported differences in behavioral strategies may be due to the different types of tasks used in different laboratories. Some researchers have reported that animals performing two-alternative choice (TAC) tasks need only 1–2 sniffs and do not increase performance with longer sampling. Others have reported that animals performing go/no-go (GNG) tasks increase sampling times and performance for difficult discriminations, arguing for temporal integration. We present results from four experiments comparing GNG and TAC tasks over several behavioral variables (e.g., performance, sampling duration). When rats know only one task, they perform better in GNG than in TAC. However, performance was not statistically different when rats learned and were tested in both tasks. Rats sample odors longer in GNG than in TAC, even when they know both tasks and perform them in the same or different sessions. Longer sampling is associated with better performance for both tasks in difficult discriminations, which supports the case for temporal integration over ≥2– 6 sniffs in both tasks. These results illustrate that generalizations from a single task about behavioral or cognitive abilities (e.g., processing, perception) do not capture the full range of complexity and can significantly impact inferences about general abilities in sensory perception.

Key words: go/no-go; odor discrimination; olfactory behavior; psychometric model; temporal integration; two-alternative choice

---

Significance Statement

Behavioral tasks and training and testing history affect measured outcomes in cognitive tests. Rats sample odors longer in a go/no-go (GNG) than in a two-alternative choice (TAC) task, performing better in GNG unless they know both tasks. Odor-sampling time is extended in both tasks when the odors to be discriminated are very similar. Rats may extend sampling time to integrate odor information up to ~0.5 s (2– 6 sniffs). Such factors as task, task parameters, and training history affect decision times and performance, making it important to use multiple tasks when making inferences about sensory or cognitive processing.

Introduction

Different laboratories often use different behavioral tasks to answer the same questions, which may confound discovery. In the olfactory literature, reports differ regarding how long rats and mice choose to sample odors in discrimination tasks and whether

they sample longer to improve performance (temporal integration). These reports used different tasks to support their differing claims.

It might be expected that temporal integration exists in olfaction, because it is present in other perceptual systems (Gold and Shadlen, 2002), but one could make arguments both for and against based upon results from different laboratories. Some studies have shown that rats and mice do not manipulate sampling times contingent on discrimination difficulty and that they need and use only 1–2 sniffs to identify an odor in a two-alternative choice (TAC) task (Uchida and Mainen, 2003; Rin-

Table 1. Odor sets

	A			B		
		Theoretical vapor pressure			Theoretical vapor pressure	
Odor set	Odor	kPa	mmHg	Odor	kPa	mmHg
Training	Anisole	0.524	3.93	Amyl acetate	0.545	4.088
Extreme	(+)-Limonene + trace	0.206	1.545	(+)-Limonene + trace	0.206	1.545
S2	Propyl proprionate	1.859	13.943	Ethyl butyrate	1.643	12.323
S3	Hexanal	1.333	9.998	Ethyl benzene	1.392	10.44
S4	Heptanone	0.23	1.725	Octanone	0.372	2.79
S5	(+)-Carvone/propyl proprionate	0.816	6.12	(+)-Limonene/propyl proprionate	0.767	5.753
S6	MIX[a] + citral	0.032	0.24	MIX[a] + citronellal	0.031	0.233
S8	Cineole	0.185	1.388	Hexyl acetate	0.219	1.643

Within each odor set, the odors are listed in order (odor A, B). Values retrieved from manufacturer websites or http://pubchem.ncbi.nlm.nih.gov/. Odors were purchased from Sigma-Aldrich (propyl proprionate, hexanal, hexyl acetate, decanol, hexanoic acid, methyl 2-fuorate), Fluka (anisole, citral), Thermo Fisher Scientific (ethylbenzene), and Acros Organics (n-amyl acetate, 2-heptanone, 2-octanone, ethyl butyrate, d-limonene, cineole, octanol, carvone).

[a]MIX: octanol; decanol; hexanoic acid; +carvone; methyl furoate.

berg et al., 2006). Even though, when forced to sample longer, mice performed better, mice did not choose longer sampling times on their own (Rinberg et al., 2006). These studies and others (Zariwala et al., 2013) suggest little integration across sniffs in practice. This does not rule out the possibility that, given the right circumstances, rodents could use longer sampling times and increase performance.

In some circumstances rats and mice do manipulate sampling times contingent on discrimination difficulty or for other reasons. Mice chose longer sampling times for some odors in a go/no-go (GNG) than in a TAC task (Abraham et al., 2004; Rinberg et al., 2006). Rats sampled longer than needed when odor exposure was unlimited (Frederick et al., 2011), and performance increased with sampling time when rats detected odors in mixtures (Rojas-Líbano and Kay, 2012). In a GNG task, mice increased sampling for hard discriminations (Abraham et al., 2004, 2012). These and other reports support temporal integration across species and phyla (Skarda and Freeman, 1987; Laurent, 1996; Friedrich and Laurent, 2004; Schaefer and Margrie, 2007).

We and others hypothesized that differing results suggest at least two ways in which rats (or mice) perform odor discriminations, mapping onto GNG and TAC tasks (Friedrich, 2006; Kay et al., 2006; Rinberg et al., 2006). Different tasks could activate different cognitive systems (e.g., hippocampal or striatal dominated), or use different optimization strategies that depend on reward, such as speed–accuracy tradeoffs. In symmetric high-reward TAC tasks (rewards delivered for all correct responses) subjects may select short sampling times even if performance is worse. In asymmetric lower-reward GNG tasks (rewards delivered only for correct go responses), subjects may be more deliberative, accumulating information about the odor over more extended time. It is possible that reliance on single behavioral tasks may have driven inconsistencies across laboratories.

Based on these apparent differences, we test the hypothesis that rats use temporal integration to improve performance with longer sampling in GNG but not in TAC tasks. This difference may be driven by reward structure. To test this hypothesis, we performed two main and two follow-up experiments coupling behavior and electrophysiology. Electrophysiological results have been presented separately and show a unified underlying sequence of oscillatory states in both tasks where the power but not the sequence of gamma (65–110 Hz) and beta (15–30 Hz) oscillations is modified by many factors, including the task (Frederick et al., 2016). We also test general task differences in several additional behavioral variables.

Behavioral analyses reported here show that rats perform better in GNG than TAC when tested on multiple odor sets only when the rats do not know both tasks. Rats sniff odors longer in GNG than in TAC regardless of whether they know one or both tasks. However, they adjust sampling times for very difficult odor discriminations in both tasks. The data show sigmoidal relationships between sampling time and performance, which argues for temporal integration in both tasks. These results reconcile past literature by establishing that reported task differences are likely not just laboratory effects; the magnitude of differences may be altered by specific task parameters and animals' training and testing histories.

Materials and Methods

Subjects

Forty-two adult male Sprague Dawley rats (Harlan; ~400–450 g throughout the experiments; RRID:RGD_550839) were used in four experiments [32 for Experiment 1 (E1); 8 for Experiments 2 (E2) and 3 (E3); and 2 for Experiment 4 (E4)]. Rats were housed in individual cages on a 14/10 h light/dark cycle (lights on at 0800 CST). All experiments were conducted during the light period. Before training, rats were dieted to 85% of their *ad libitum* weight and maintained at this level for the remainder of the experiments. All methods and procedures were done under veterinary supervision and oversight of the University of Chicago Institutional Animal Care and Use Committee in accordance with Association for Assessment and Accreditation of Laboratory Animal Care standards.

Odor sets

We used nine odor sets with different and overlapping subsets for the different experiments (Table 1; Fig. 1). Each odor set consisted of two odors (labeled A or B). Each odor was either a monomolecular odorant [training odor set (Training), extreme odor set (Extreme), and odor sets 2–4 (S2–S4) and 8 (S8)] or a mixture of odorants [odor sets 5 (S5) and 6 (S6)]. For GNG, odor A was the "go" signal and odor B was the "no-go" signal. For TAC, odor A was the "go left" signal and odor B was the "go right" signal. Our prior research showed that mapping of behavioral response (go, no-go, go-left, go-right) to odor (A or B) did not selectively affect sampling or performance; therefore, we did not randomize assignment of behavior response to odor A or B (Frederick et al., 2011).

The extreme odor set was composed of (+)-limonene for both odors plus trace containments. Each test tube, which held the liquid odorant, was placed in a different experimental room, without a stopper, for ≥4 h. The extreme odor set was remade approximately every 2 weeks. This odor set was inspired by a report that rats could discriminate between mineral oil made by different companies (Gamble and Smith, 2009) and was designed to be the most difficult for rats to discriminate.

Behavior protocols

General. For all operant tasks, rats were trained and tested in one of two operant box designs. Our general training paradigms and box designs
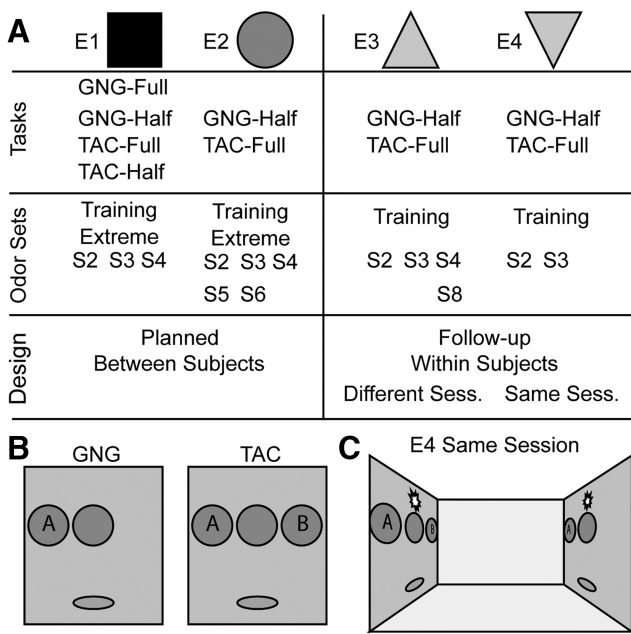
Figure 1. Experiment design. ***A***, Each column lists the type of behavior, odor sets, and design used for each experiment. The symbols next to each experiment are those used in the following figures. The odor sets and odors used are identified in Table 1. E1 and E2 were between-subjects designs (rats learned 1 task). E3 and E4 were within-subjects designs (rats learned both tasks). ***B***, Schematic of the behavioral panels used. For both tasks, there is a center odor port. When rats stick their noses into the port, odor is delivered for as long as they remain in the port. For GNG, there is a single response port (A) in which a rat makes a response to odor A. No response is made to odor B. For TAC, there are two response ports. The left port is for responses to odor A and the right port for responses to odor B. See Materials and Methods for complete details. ***C***, Schematic of the operant box used in E4 in which rats performed GNG and TAC tasks in the same session. On one side of the box was a GNG panel and on the other side of the box was a TAC panel. A light was illuminated on the panel to indicate which task to do for any given trial. See Materials and Methods for complete details.

have been documented previously (Frederick et al., 2011; Rojas-Líbano and Kay, 2012). Briefly, the box was made from Med Associates parts and controlled by their MedPC-IV software. The front of the box contained an odor port (located in the center), one or two response ports (located to either side of the odor port), and a reward dish (located below the odor port). A house light was attached to the back panel.

During training and testing, rats nose-poked into a central odor port in which they received an odor stream (~14% concentration, ~1.2 l/min). The delay to odor arrival is ~60 ms, as calculated from airflow, tubing length, and solenoid closing time. In practice, this delay could be longer, but we assume the variance in delay is stable over the randomized ordering of odor tests across the rats. The odor stayed on for as long as the rat remained within the port (i.e., there was no upper limit on how long rats could sample). The longest single sampling bout observed was 18.9 s (Fig. 2a).

We used our standard controls (Frederick et al., 2011; Rojas-Líbano and Kay, 2012) to ensure that rats were not able to detect odors early and that they were not able to discriminate based upon nonolfactory cues.

For all experiments, rats were trained in three phases using our standard training protocols (Frederick et al., 2011). In Phase 1, rats learned to nosepoke in the odor port. In Phase 2, rats learned to respond via a nosepoke into a response port (located to the left of the odor port) after first nosepoking in the odor port. Training was the same for all rats in Phases 1 and 2, but differed in Phase 3. For GNG training, rats learned to withhold responding (i.e., nosepoking into a response port) for a new odor (B). For TAC training, rats learned to respond to a new odor (B) by nosepoking into a (new) right response port. Errors for incorrect go responses in either task were penalized by an immediate 7 s lights-off timeout period that was added to the normal intertrial interval period. A reward was one 45 mg sucrose pellet (Bio-Serve Dustless Precision Pellets

#F06233), delivered automatically after a correct response for a rewarded trial.

Timeline of experiments and rationale. The details of each experiment are described below. Here, we provide a brief overview of how and why the studies were included in this paper. We originally had designed and planned for only E1 and E2. E1 had the largest sample size because it involved only behavior and was designed to test task and reward rate effects. E2 was a replication of E1 with the addition of electrophysiological recordings. As such, we were limited in the number of animals used. Therefore, E2 has a smaller sample size. E3 and E4 were designed after E1 was finished but E2 was still being conducted. E3 and E4 were initially planned as exploratory experiments for future research. Because the results of these two experiments complicate the results from E1 and E2, we have included these initial results in this report so as to present the most accurate picture of the underlying behaviors and systems as is currently available from our data.

Note on sample sizes. Sample sizes in E1 and E2 were based on common numbers in previous experiments in the literature and what was feasible for the laboratory. Given the complexity of the experiments and behaviors involved, both experiments are likely underpowered, which is a chronic problem in cognitive neuroscience (Szucs and Ioannidis, 2016).

Because E3 and E4 were intended as a proof-of-concept test of the results from E1 and E2, the numbers are small in each. However, within-subjects comparison of the tasks provides a powerful illustration and complication to the comparisons seen in E1 and E2.

E1. Rats learned one of four behaviors (*n* = 8/group) that resulted from crossing task (GNG, TAC) with reward (Full, Half). Rats in the GNG-Full group received a reward if they were correct for either the go or the no-go trials. Rats in the GNG-Half condition (i.e., standard GNG) only received a reward for correct go trials. Rats in the TAC-Full condition (i.e., standard TAC) received a reward for correct left or right trials. Rats in the TAC-Half condition received a reward with a 50% probability after a correct response for either stimulus. The TAC-Half condition was intended to mimic the total reward probability of the standard GNG (i.e., GNG-HR) task, and the GNG-Full condition was intended to mimic the total reward probability of the standard TAC task. Training for TAC-Half had an additional step. Rats were initially trained with full reward on the training odor set and then the reward level was decreased over a few days. After transfer from training to test odor sets, reward probabilities remained at 50%. Following training, rats were tested on four odor sets (Extreme, S2, S3, and S4) for 3 d each. The odor set order was randomized for each rat. Rats were run for 300 attempted trials (i.e., a trial in which a rat nosepoked into the odor port) or 400 total trials, whichever came first.

E2. This experiment was designed as a replication of E1 with the addition of local field potential (LFP) recordings from the olfactory bulb and pyriform cortex. LFP results are the focus of a separate paper (Frederick et al., 2016). Two groups (*n* = 4/group) of rats were trained in either standard GNG (GNG-Half) or standard TAC (TAC-Full) using the same procedures as E1. Rats were tested on new odor sets until they reached 2 consecutive days >80% or reached 10 d on the same odor set (a total of 349 sessions across 8 rats). Due to technical issues, six times across four subjects, rats were run more than the maximum number of 10 d on the same odor set. We noted no significant differences in results between including all days or not. We have chosen to retain all data. In E2, rats were run for 150 attempted trials or 200 total trials. We decreased the number of trials because the animals in this experiment were also implanted with electrodes. We wanted to minimize the wear on the implants and fatigue for the tethered rats. In our experience, implanted rats are often less likely to complete >200 trials per session.

E3. E3 was a follow-up experiment performed at the conclusion of E2; it was designed to test whether the observed task differences seen in E1 and E2 would remain within subjects that had learned both tasks. Following completion of E2, the remaining three rats were trained on the behavior that they had not previously learned with the training odor set. Two rats that had previously been tested in TAC learned GNG and one rat that had previously been tested in GNG learned TAC. Once the rats learned the new behavior, we taught them to perform both behaviors in the same day in separate sessions, with the order of the two behaviors
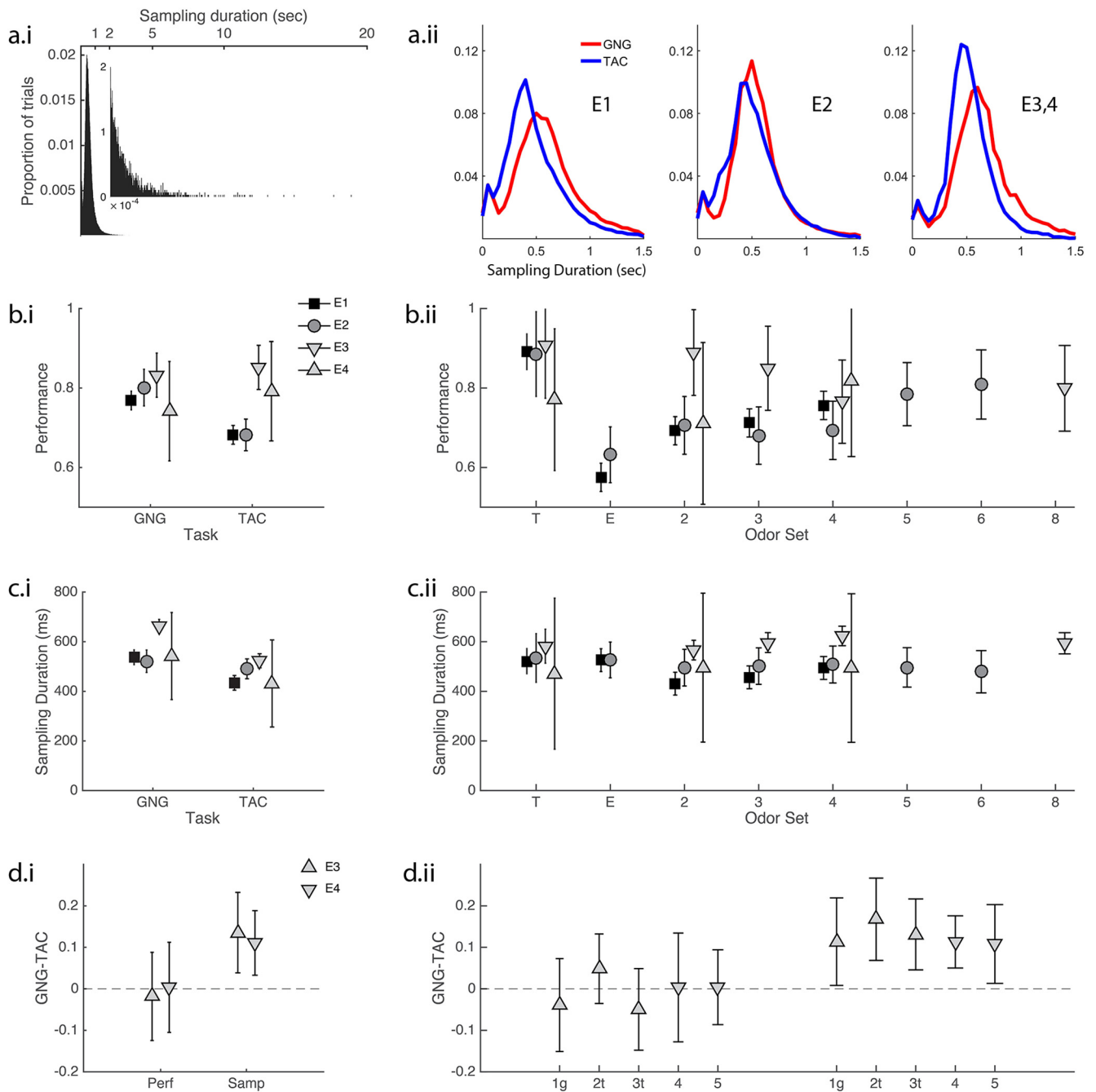
Figure 2. Performance and sampling duration. Summary statistics are in the results. *a*, Sampling duration distributions. *a.i*, All experiments, odors and rats combined. The background figure contains the entire distribution, and the inset shows the tail from 2 to 20 s. Note that there are many trials in which rats sniff for >5 s. *a.ii*, 0 – 1.5 s sampling time distributions for GNG and TAC. Experiments as indicated on the plots. *b*, Performance means and CIs from the GLS model. *b.i*, Performance by task grouped by experiment. For E1 and E2, GNG rats perform better than TAC rats. For E3 and E4, there is no performance difference. *b.ii*, Performance by odor set grouped by experiment. See results for multiple-comparison statistics. See Table 2 for means and CIs from GLS model fits. *c*, Sampling duration results. *c.i*, Sampling duration by task and experiment. For E1, E3, and E4, GNG rats sampled significantly longer than TAC rats. The difference was not significant in E2. *c.ii*, Sampling duration by odor set and experiment. *d*, E3 and E4 individual subject results. *d.i*, Within-subject E3 and E4 mean differences (GNG − TAC) in performance and sampling duration. *d.ii*, Individual subject differences (GNG − TAC) in performance and sampling duration in E3 and E4. Rats 1–3 are from E3; rats 4 and 5 are from E4. "g" indicates E3 rat was GNG in E2. "t" indicates rat was TAC in E2. Values are mean ± SD.

randomly assigned each day. After achieving >80% on the training odor set, they were retested on odor sets S2–S4 and a novel odor set, S8. Rats were run for 4 d per odor set. Two rats had odor sets S2–S4 interleaved with a different order for each rat, so that they were tested on a different odor set each day. The third rat was tested in block fashion on odor sets S2–S4, as described for E1 and E2, and then S8. Due to technical issues, this rat was run for 6 d on S2 and for 2 d on S8. Rats achieved >80% performance. As in E2, rats were run for 150 attempted trials or 200 total

trials per session. The LFP results from this experiment are reported in a separate paper (Frederick et al., 2016).

E4. E4 was a second follow-up within-subjects experiment designed to test whether the task differences we observed in E3 would occur when a rat performed both behaviors in the same session. We initially trained rats on TAC as in E1 and E2. Once they learned the behavior, we trained them on GNG on the opposite side of the same operant chamber. Following learning, we trained them to perform both behaviors in the same

session with trial type randomly interleaved and signaled by lights on each side of the cage. After training, we tested the rats on odor sets S2 (for 3 d) and S4 (for 4 d). Rats were run for 150 attempted trials or 200 total trials (the training for this experiment takes many more weeks than for the other experiments).

Statistical methods

For each session, we computed several session statistics that formed the basis for further statistical analyses. We calculated the sampling duration (and other delay statistics) as the session median to account for skewed distributions. Performance was calculated as the weighted average of individual odor performance $\left(P_w = \dfrac{P_A + P_B}{2} \right)$. We did this to ensure that the session performance statistic was not biased, given that rats did not always receive exactly 50% of each odor type (odors were randomly selected on each trial, with replacement). In GNG sessions, rats tended to show ceiling-level performance for the go odor. For example, if a rat performing GNG chose the go behavior on every trial and did 100 A trials at 100% and 50 B trials at 0%, the simple average performance would be 67% (100 of 150), whereas the weighted performance would be 50%, which is, we argue, a better representation of the actual session performance to be able to compare tasks. An alternative to this would have been to use signal detection theory's d' (Stanislaw and Todorov, 1999). In general, results did not differ if d' was used.

For response delay, we used all responses (correct and incorrect) for the session average.

Generalized least-squares statistical models. For all analyses, we used generalized least-squares regression (GLS), which is a generalization of ordinary least-squares regression and a part of a larger group of general Gaussian methods that allow for modeling data with correlated terms (Clifford and McCullagh, 2006; McCullagh and Clifford, 2006). Analysis was done in R using the gls function from the nlme package (Pinheiro and Bates, 2000; Pinheiro et al., 2013). *P* values for main effects (i.e., fixed effects) were calculated using the Wald *F* test. Ninety-five percent confidence intervals (95% CIs) were computed using lsmeans function (or intervals function) and are reported in the text using the following general form: Variable = [Lower, Upper]. The general form of the GLS models was as follows: DV ~ task + odorSet + CoV.

Task (e.g., GNG, TAC) and odor set (e.g., Training, Extreme, S2–S6, S8) were nominal variables. CoV was an exponential covariance matrix meant to control for correlated data terms.

For E1, the full model was as follows: DV ~ taskType * rewardType + odorSet + CoV, where taskType (i.e., task) was GNG or TAC and rewardType (i.e., reward) was Full or Half. If the task–reward interaction was not significant, we removed the interaction from the statistical model and recomputed it. If, in the new statistical model, reward was not significant, we removed it and used the resulting statistical model (i.e., task + odor set). We did this for conciseness and to use as similar a model for all experiments as possible.

For *post hoc* comparisons, we used a Bonferroni correction and report significant results in the text based on the adjusted α level.

We also tested to see whether rats were biased in their responses to either odor A or odor B. To determine whether there was a significant bias, we first computed a difference value of odor A − odor B. The difference value was used as the DV to build GLS models as described above. From these models, we computed 95% CIs using lsmeans for task. We then used these 95% CIs to test whether there is a significant bias present. We consider there to be a significant bias present if 0 falls outside of the computed CIs. We present the 95% CIs used to make these judgments in the text.

When we write "no differences" in the text, or write that an effect is not present or that a statistical model was not able to detect an effect, we do not mean to accept the null hypothesis. Instead, we mean merely to say that the statistical model did not find a significant effect based upon the conventional use of 0.05 as a threshold for significance. In general, we do not have sufficient power to be able to accept the null hypothesis in any of our analyses in this set of experiments. The lack of a significant effect could be because (1) there really is no effect or (2) we did not have sufficient power to detect the effect given our sample sizes.

Conditional probability plots. We computed empirical conditional probability plots to show the trial-level relationship between sampling duration and performance. We created the plots by pooling trials from task and odors together (i.e., GNG-A, GNG-B, TAC-A, and TAC-B). Given that our hardware time resolution was 10 ms, we used 10 ms bins. For each bin from 10 to 1000 ms we calculated the number of attempted trials and the number of correct trials for that bin and used these values to compute the conditional probability of a trial being correct given a sampling duration as well as the binomial CIs around that probability.

To test whether there was an association between session level (median) sampling duration and performance, we used GLS models. For these statistical models, we added sampling duration as a covariate. We also checked for possible interactions between task, odor set, and sampling duration.

Results

Sampling duration and its relationship to performance are key measures in evaluating olfactory behavior. When subjects sample for brief periods and do not increase sampling for difficult discriminations, investigators conclude that they do not use temporal integration, or the related speed–accuracy tradeoff, in odor discrimination (Rinberg et al., 2006; Zariwala et al., 2013). However, given that there are reported instances in which mice and rats can and/or do sample longer and perform better (Rinberg et al., 2006; Frederick et al., 2011; Rojas-Líbano and Kay, 2012) and do increase sampling for difficult odor discriminations (Abraham et al., 2004), it is important to determine what factors might drive these different results across laboratories. Are these differences in the literature due to laboratory effects or to task differences, and do rats adjust sampling times for difficult discriminations and/or appear to use temporal integration in one or both types of tasks?

To test whether reported differences in the olfactory literature are a result of the particular tasks (GNG and TAC) being used, we trained and tested rats in these two tasks using between-subjects (E1 and E2) and within subjects (E3 and E4) designs. We tested whether tasks were different across several behavioral outcome variables: performance, sampling duration, response delay, start delay, and reward interval (i.e., time between rewards). In addition to these measures, we tested the association between sampling duration and performance and examined conditional probability plots. Based on previous research, which suggested that the disagreements in the literature may lie in the different tasks chosen by different laboratories, we expected that rats performing a GNG task would sample odors longer and perform better than in TAC. We also expected that given the long sampling times shown at the end of training and our previous results in an odor-background task, we might find that rats do better when sampling longer, supporting temporal integration, at least in the GNG task.

We tested the influence of overall reward probability within and across tasks in E1. Reward condition (Full or Half) and interactions with tasks did not produce any significant effects on performance or sampling duration. We therefore omit reporting of the reward conditions in E1, except where they were significant for other factors.

The results of GLS models are presented in Tables 2–5, cited where relevant in the text. When discussing significant effects in the text, we report the *F* statistics and associated *p* values along with 95% CIs, if they are not depicted in an associated figure. For nonsignificant effects, we write $p > 0.05$ for single effects and group multiple nonsignificant effects together as $ps > 0.05$. We present 95% CIs for task effects, even if the effect was not significant, for comparative purposes.

Table 2. GLS model results for performance and sampling duration

	Factors	Levels	E1	E2	E3	E4
Performance		Intercept	0.93 (0.02)***	0.95 (0.06)***	0.90 (0.07)***	0.77 (0.11)***
	Task^{a,b}	GNG	Reference	Reference	Reference	Reference
		TAC	−0.09 (0.02)***	−0.12 (0.03)***	0.02 (0.02)	−0.004 (0.06)
	Odor set^{a,b}	Train	Reference	Reference	Reference	Reference
		Extreme	−0.32 (0.03)***	−0.25 (0.06)***	—	—
		S2	−0.20 (0.03)***	−0.18 (0.07)**	−0.02 (0.09)	−0.08 (0.14)
		S3	−0.18 (0.03)***	−0.21 (0.07)**	−0.06 (0.09)	—
		S4	−0.14 (0.03)***	−0.19 (0.07)**	−0.14 (0.08)	0.003 (0.14)
		S5	—	−0.10 (0.07)	—	—
		S6	—	−0.08 (0.07)	—	—
		S8	—	—	−0.11 (0.09)	—
Sampling duration		Intercept	573 (28)***	550 (52)***	650 (40)***	527 (149)***
	Task^{a,c,d}	GNG	Reference	Reference	Reference	Reference
		TAC	−103 (21)***	−31 (30)	−135 (16)***	−111 (22)***
	Odor set^{a}	Train	Reference	Reference	Reference	Reference
		Extreme	4 (35)	−8 (62)	—	—
		S2	−91 (35)**	−39 (62)	−18 (49)	24 (209)
		S3	−65 (35)	−33 (62)	18 (49)	—
		S4	−27 (35)	−26 (63)	39 (49)	23 (208)
		S5	—	−38(64)	—	—
		S6	—	−55 (66)	—	—
		S8	—	—	7 (50)	—

Statistical significance is based upon Wald F tests, which are indicated in superscripts next to each factor (e.g., Task, Odor set) with values (a, E1; b, E2; c, E3; d, E4) for Tables 2–5. Odor sets are identified in the coefficient level column by name or S + number (See Table 1 for odors). Significant coefficient based upon t value given as follows: *0.05, **0.01, ***0.001. All durations are rounded to the nearest millisecond. Values in parentheses are SEs. Blank cells indicate factors or odor sets that were not used in a given experiment. The intercept represents the expected value given the reference level for each factor. A significant intercept means that it is statistically different from zero. For example, the intercept in E1 performance represents expected performance for a GNG rat on the training odor set and is different from zero. Values for levels other than the reference are with respect to the factor reference. To get the expected value of any specific combination of levels, one simply adds the value for the chosen level (within a factor) to the intercept. So, for E1 we would expect that a rat engaged in a TAC task would perform 9% worse than a GNG rat, on average, across all odor sets; or, that in E1 a GNG rat would perform, on average, the extreme odor set 32% worse than the training odor set.

Table 3. GLS model results for response delay E1 in milliseconds

Factors	Levels	E1
	Intercept	403 (41)***
Task^{a}	Task GNG	Reference
	Task TAC	87 (39)*
Reward^{a}	Reward Full	Reference
	Reward Half	92 (39)*
Task × Reward^{a}	(= TAC, half)	−172 (55)**
Odor set	Train	Reference
	Extreme	−15 (45)
	S2	19 (45)
	S3	34 (45)
	S4	21 (45)

Interpretation follows as outlined in Table 1. There is a significant interaction in E1. The notation in the level column for the interaction means to add this value to the model when calculating the expected value of a rat engaged in TAC-Half task (e.g., a TAC-Half rat on odor set S3 would have an expected response delay of 403 + 87 + 92 − 172 + 34 = 444 ms).

Table 4. GLS model results for start delay in milliseconds

Factors	Levels	E1	E2	E3	E4
	Intercept	809 (48)***	440 (93)***	495 (134)***	1517 (198)***
Task^{d}	GNG	Reference	Reference	Reference	Reference
	TAC	−199 (43)***	−81 (44)	−40 (46)	−486 (67)***
Reward^{a}	Reward Full	Reference	Reference	Reference	Reference
	Reward Half	−262 (43)***	—	—	—
Task × Reward^{a}	(= TAC-Half)	229 (60)***	—	—	—
Odor set	Train	Reference	Reference	Reference	Reference
	Extreme	−156 (52)**	−124 (103)	—	—
	S2	−193 (52)***	−15 (104)	−156 (171)	−135 (264)
	S3	−187 (52)***	−6 (104)	−239 (170)	—
	S4	−82 (52)	−77 (105)	−93 (170)	−14 (262)
	S5	—	−84 (107)	—	—
	S6	—	−50 (112)	—	—
	S8	—	—	−161 (171)	—

Blank cells indicate factors or odor sets that were not used in a given experiment. Interpretation follows as outlined in Table 1. There is a significant interaction in E1. The notation in the level column means to use this value when calculating the expected value of a rat engaged in TAC-Half task.

Performance

There was a marked difference in performance dependent on whether rats were trained and tested on only one task (E1, E2, between subjects) or on both tasks (E3, E4, within subjects). In E1 and E2, rats tested in GNG outperformed rats tested in TAC by 9–12%, depending upon the experiment (Fig. 2b.i; Table 2; E1: $F_{(1,404)} = 26.32, p < 0.0001$; E2: $F_{(1,286)} = 16.61, p = 0.0001$). In E3 and E4, rats performed both tasks either in separate sessions (E3) or in the same session (E4), and we detected no significant differences in performance between the tasks (Fig. 2b; Table 2; $ps > 0.05$).

We also found a significant effect of odor set on performance in E1 and E2, but not in E3 and E4 (Fig. 2b.ii; Table 2; E1: $F_{(4,404)} = 31.55, p < 0.0001$; E2: $F_{(6,286)} = 3.97, p = 0.0008$; E3, E4: $ps > 0.05$). In E1, rats performed the best on the training odor set ($ps < 0.0001$) and worst on the extreme odor set ($ps < 0.0001$). In E2, the significant contrasts were more mixed. E2 rats performed better on the training odor set than the extreme odor

set or S3 ($ps < 0.002$). They also performed better on S6 than the extreme odor set ($p = 0.002$). See Figure 2b for all 95% CIs.

When performing the GNG task, rats exhibited a performance bias, showing better performance for odor A (go odor; performance difference A − B: E1 = [0.31, 0.46]; E2 = [0.25, 0.45]; E3 = [0.17, 0.36]; E4 = [0.29, 0.70]; we consider significant the 95% CIs that do not include zero). In the GNG task, rats tend to choose the go-response as a default response, even when they do not remain in the odor port long enough to receive a full sniff of the odor (<100 ms). Rats tested in TAC did not show significant A − B biases (E1 = [−0.07, 0.08]; E2 = [−0.08, 0.10]; E3 = [−0.04, 0.15]; E4 = [−0.14, 0.28]).

In summary, rats trained in GNG perform better than those trained in TAC, but only when the rats know only one task (E1

Table 5. GLS model results for reward interval in seconds

Factor	Level	E1	E2	E3
	Intercept	34.27 (3.57)***	25.69 (3.08)***	34.66 (24.62)
Task[d]	GNG	Reference	Reference	Reference
	TAC	−6.68 (2.82)*	−6.06 (1.51)***	−27.28 (10.95)*
Reward[a]	Reward Full	Reference	—	—
	Reward Half	−0.45 (2.80)	—	—
Task × Reward[a]	(= TAC-Half)	23.23 (3.97)***	—	—
Odor set	Train	Reference	Reference	Reference
	Extreme	3.20 (3.77)	6.19 (3.45)	—
	S2	−3.25 (3.77)	7.44 (3.48)*	1.65 (28.91)
	S3	−1.93 (3.76)	7.71 (3.47)*	2.28 (28.61)
	S4	−0.82 (3.77)	7.04 (3.49)*	32.76 (28.41)
	S5	—	3.35 (3.59)	—
	S6	—	2.98 (3.73)	—
	S8	—	—	3.76 (28.91)

Blank cells indicate factors or odor sets that were not used in a given experiment. Interpretation follows as outlined in Table 1. There is a significant interaction in E1. The notation in the level column means to use this value when calculating the expected value of a rat engaged in TAC-Half task.

and E2). When the rats know both tasks (E3 and E4), rats perform both tasks equally well. As might be expected, some odor sets are easier to discriminate than others, as evidenced by higher performance, and this is true for both tasks.

Sampling duration
Rats showed a large distribution of odor-sampling times (range: 10 ms to 18.9 s; Fig. 2a). Rats often engaged in very long sniffing bouts, represented by the long and well populated tail on the right side of the distribution. These long bouts appear unique to our setup; we do not cap the maximum odor stimulus delivery time; the odor stays on as long as the rat stays in the odor port and only that long.

Sampling duration showed the most consistent difference between tasks. In all experiments, rats tested in GNG sampled longer than rats tested in TAC. Differences were in the same direction in all four experiments and statistically significant in E1, E3, and E4 (Fig. 2c; Table 2; E1: $F_{(1,404)} = 23.35$, $p < 0.0001$; E3: $F_{(1,86)} = 72.56$, $p < 0.0001$; E4: $F_{(1,28)} = 26.10$, $p < 0.0001$), but not in E2 ($p > 0.05$).

Rats sampled different odor sets for different lengths of time in E1, but not in the other experiments (Table 2; E1: $F_{(4,404)} = 3.00$, $p = 0.0184$; E2–E4: $ps > 0.05$). *Post hoc* multiple comparisons show that only the extreme odor set and S2 were significantly different with rats sampling Extreme 95 ms (~1 sniff) longer than S2 ($p = 0.0044$).

Rats tested in GNG showed a sampling bias in E1–E3, with longer sampling times for odor A in E1, longer sampling for odor B in E2 and E3, and no bias in E4 (A − B: E1 = [8, 49 ms]; E2 = [−94, −45 ms]; E3 = [−85, −22 ms]; E4 = [−45, 37 ms]). Rats tested in TAC showed a bias for sampling odor A longer in E2 and E3 but no bias in E1 and E4 (A − B: E1 = [−37, 4 ms]; E2 = [14, 55 ms]; E3 = [20, 84 ms]; E4 = [−40, 42 ms]). Therefore, we conclude that sampling biases, where they exist, may not be driven by a particular subsequent behavioral response (go, no-go, go-left, go-right).

In general, rats sample odors longer in the GNG than in the TAC context, even when they know both tasks. There is also some evidence that the difficulty of the odor discrimination can be associated with longer sampling times in both tasks.

Within-subjects results on sampling duration and performance
E3 and E4 together provide a powerful assessment of task effects within subjects who perform both tasks. In E3, the rats performed

GNG and TAC tasks in separate sessions randomly in the morning and afternoon of each day. In E4, the rats performed both tasks randomly cued within the chamber on each trial. Each of these rats presented a significant sampling difference such that GNG > TAC (Fig. 2d). Performance for each rat was not different across the two tasks. It made no difference in these measures whether the rats performed the two tasks in the same or in different sessions.

Performance and sampling duration association
In addition to the above planned analyses, we also examined whether there was an association between the session sampling duration and session performance. We used GLS statistical modeling as described in Materials and Methods.

For E1, a full interaction model that included interactions between task, reward type, odor set, and sampling duration as well as a simplified main effects-only model (i.e., it did not have interaction terms), did not show any significant associations between sampling duration and performance at the level of the sessions ($ps > 0.05$).

For E2, a full interaction model (i.e., with interactions between task, odor set, and sampling duration) found a main effect of sampling duration ($p < 0.05$), but not of the interaction terms ($ps > 0.05$). We therefore removed the interaction terms and computed a simpler main effects-only model wherein sampling duration remained significant ($F_{(1,285)} = 12.72$; $p = 0.0004$; sampling duration coefficient, 0.025 performance/100 ms; SE = 0.007 performance/100 ms), which means that the rats increased performance by 2.5% with each 100 ms of sampling time.

For E3, the full interaction model (i.e., with interactions between task, odor set, and sampling duration) showed a significant interaction between task and sampling duration (interaction term TAC × Sampling Duration: $F_{(1,72)} = 4.59$, $p = 0.0355$; coefficient, 0.20 performance/100 ms; SE = 0.48 performance/100 ms). This interaction term means that, everything else being equal, rats performing TAC have a larger performance–sampling duration association than when performing GNG. However, the simpler main effects-only model produced a better fit than the interaction one as determined by comparison of the Bayesian information criterion (BIC; interaction model BIC, −46.94; main effects model BIC, −84.89; smaller BIC is better and a difference between BIC values >10 is often seen as very strong evidence against the model with a larger BIC value). That the simpler main effects-only model was a better fit than the interaction model may indicate that the significant interaction term found is a false positive. Alternatively, the interaction effect may be real but our sample size too small to reliably detect the effect.

For E4, there were no significant interactions with sampling duration. There was a main effect of sampling duration in the interaction model. A simple main effects-only model retained the significant effect ($F_{(1,27)} = 16.65$; $p = 0.0004$; coefficient, 0.065 performance/100 ms; SE = 0.016 performance/100 ms).

In summary, there were several significant associations between session-level sampling duration and performance, where increased sampling durations were associated with increased session performance. This effect ranged from 2.5 to 6.5% increase in correct performance for each 100 ms of odor sampling (~1 sniff at 10 Hz sampling frequency). The effect is variable across experiments, suggesting that there is either too much noise due to sample size or that the effect is contingent on task parameters. Still, the results suggest that rats that sample longer tend to perform better.

Response delay

Response time becomes part of the decision time in some experiments, particularly when the response does not require the subject to exit the odor port (Abraham et al., 2004, 2012; Doucette and Restrepo, 2008). With our design, we at least partially separated the odor sampling and the motor portion of response times (it is likely that a portion of the odor-sampling period contains motor preparation; Frederick et al., 2016). Response delay is the time from the end of sampling until a response was made into one of the response ports. The median values did not vary much within an experiment, but across experiments there are some differences. Only E1 showed significant effects of any of the factors. E2–E4 effects were not significant ($ps > 0.05$; E2: GNG = [389, 609 ms]; TAC = [255, 459 ms]; E3: GNG = [270, 314 ms]; TAC = [268, 312 ms]; E4: GNG = [271, 430 ms]; TAC = [280, 439 ms]). The single significant effect was a task–reward-type interaction in E1 [$F_{(1,402)} = 9.85$; $p = 0.0018$; GNG-Full = [361, 469 ms]; GNG-Half = [454, 561 ms]; TAC-Full = [448, 556 ms]; TAC-Half = [369, 476 ms]) with main effects of task, reward, and odor set not significant ($ps > 0.05$). In the full-reward condition, GNG rats appear to respond more quickly than TAC rats, and in the half-reward condition, the opposite relationship appears. However, Bonferroni-corrected *post hoc* comparisons resulted in none of the comparisons being below the corrected significance level ($\alpha = 0.0083$, $ps > 0.017$). We do not address biases in response time, because the GNG task does not have two equal types of responses to compare to the TAC task.

In summary, we found a significant interaction of task and reward type in E1. However, Bonferroni-corrected *post hoc* comparisons were not significant, which may indicate that the effect is a false positive or that we lacked sufficient power to adequately detect differences in the comparisons. We did not find any significant effects on response delay in the other experiments.

Start delay

The amount of time a rat takes to initiate a trial after the cue light is illuminated gives us an indication of how motivated the subject is to perform the behavior. There were significant effects of task on start delay in E1 and E4 (Table 4; E1: $F_{(1,402)} = 8.01$, $p = 0.0049$; E4: $F_{(1,28)} = 52.13$, $p < 0.0001$). In E1 there was also a significant task–reward interaction as well as a significant effect of reward on start delay (Task × Reward: $F_{(1,402)} = 14.45$, $p = 0.0002$; Reward: $F_{(1,402)} = 24.17$, $p < 0.0001$). In E1, GNG-Full rats took the longest time to initiate a trial once the light was illuminated ($P < 0.0001$; GNG-Full = [626, 744 ms]; GNG-Half = [363, 482 ms]; TAC-Full = [426, 545 ms]; TAC-Half = [393, 512 ms]). These numbers suggest that TAC rats' start-delay values did not vary by reward condition, while GNG rats' start delays were decreased in the Half relative to the Full reward condition. In E4, GNG start delay was nearly ~0.5 s longer than TAC, even though the standard GNG and TAC tasks are Half and Full reward, respectively (GNG = [1239, 1695 ms]; TAC = [754, 1209 ms]). However, E4 was different from each of the other experiments in that the tasks were randomly interleaved at the trial level. Many factors other than the task context may affect a rat's motivation to start a trial in this condition. The odor set influenced the start delay only in E1 ($F_{(4,402)} = 5.05$, $p = 0.0006$), with the training set producing longer delays than the extreme odor set, S2, and S3 ($ps < 0.0026$; Training = [556, 714 ms]; Extreme = [415, 543 ms]; S2 = [378, 506 ms]; S3 = [384, 512 ms]; S4 = [488, 617 ms]).

There were no effects on start delay in E2 and E3 ($ps > 0.05$; E2: GNG = [320, 458 ms]; TAC = [251, 365 ms]; E3: GNG = [255, 476 ms]; TAC = [215, 436 ms]).

In summary, we found significant effects of task on start delay in E1 and E4 with rats performing GNG taking longer to initiate a trial than rats performing TAC. There was a significant task–reward interaction in E1 that showed that GNG-Full rats took the longest to start a trial.

Reward interval

We define the reward interval as the time between successive rewards and use this measure to inform our reasoning about the reward outcome of a behavioral strategy over a session. In general, we expect that if rats take the same amount of time to complete the same number of trials in all tasks, the reward rate should be determined by task (e.g., GNG-Half < TAC-Full) and by performance because of missed rewards and penalty delays. Because rats do not perform the same number of trials in the same amount of time, and skip some trials, the reward interval should be affected by task, performance, and motivational factors. We found significant task differences in reward intervals in E1–E3 (Table 5).

In E1, we found a significant task–reward interaction (Task × Reward: $F_{(1,402)} = 34.15$, $p < 0.0001$). TAC-Half rats had the longest reward interval, and TAC rats differed across reward conditions but GNG rats did not ($ps < 0.0001$; GNG-Full = [29.76, 37.65 s]; GNG-Half = [29.31, 37.20 s]; TAC-Full = [23.03, 31.02 s]; TAC-Half = [45.85, 53.77 s]). Also in E1, there was a significant main effect of task; the GNG reward interval was shorter than TAC, likely driven by better GNG performance ($F_{(1,402)} = 6.42$; $p = 0.0116$; GNG = [30.66, 36.31 s]; TAC = [35.57, 41.27 s]). As expected, the time between rewards for the half-reward tasks was longer than for the full-reward tasks ($F_{(1,402)} = 30.99$; $p < 0.0001$; Full = [27.52, 33.21 s]; Half = [38.70, 44.36 s]).

In E2 and E3, rats tested in TAC had a shorter reward interval than those tested in GNG (E2: $F_{(1,286)} = 14.98$, $p = 0.0001$; GNG = [28.30, 32.99 s]; TAC = [22.61, 26.56 s]; E3: $F_{(1,86)} = 6.21$, $p = 0.0146$; GNG = [23.60, 61.89 s]; TAC [−3.68, 34.61 s]). This effect may be attributed in large part to the reward differences in the two tasks in E2 and E3. There were no effects of odor set in E2 and E3 ($ps > 0.05$).

We could not test reward interval in E4 because of the randomly interleaved trial structure. The mean (SD) for reward interval was 33.74 (8.50) s. We calculated this as the session duration divided by the number of correct rewarded trials. For GNG, correct rewarded trials are just correct odor A trials. For TAC, both odors A and B were rewarded for correct responses.

In summary, the reward interval varied dependent on task and reward type with significant interactions in E1. TAC rats varied in reward interval across the Full and Half reward conditions, but GNG rats did not. These differences are a combination of performance differences between GNG and TAC, reward schedule, and likely other unknown factors.

Relationship between performance and sampling duration

A basic question asked in the olfactory literature is: "If rats (or mice) sniff longer, do they do better?" This implies at least two possible observations. First, do rats that sniff longer, on average, in a session perform better than rats that sniff shorter (see performance and sampling duration results above)? Second, within a session at the trial level, are rats more likely to get a given trial correct if they sniff longer (e.g., is there a sigmoidal relationship
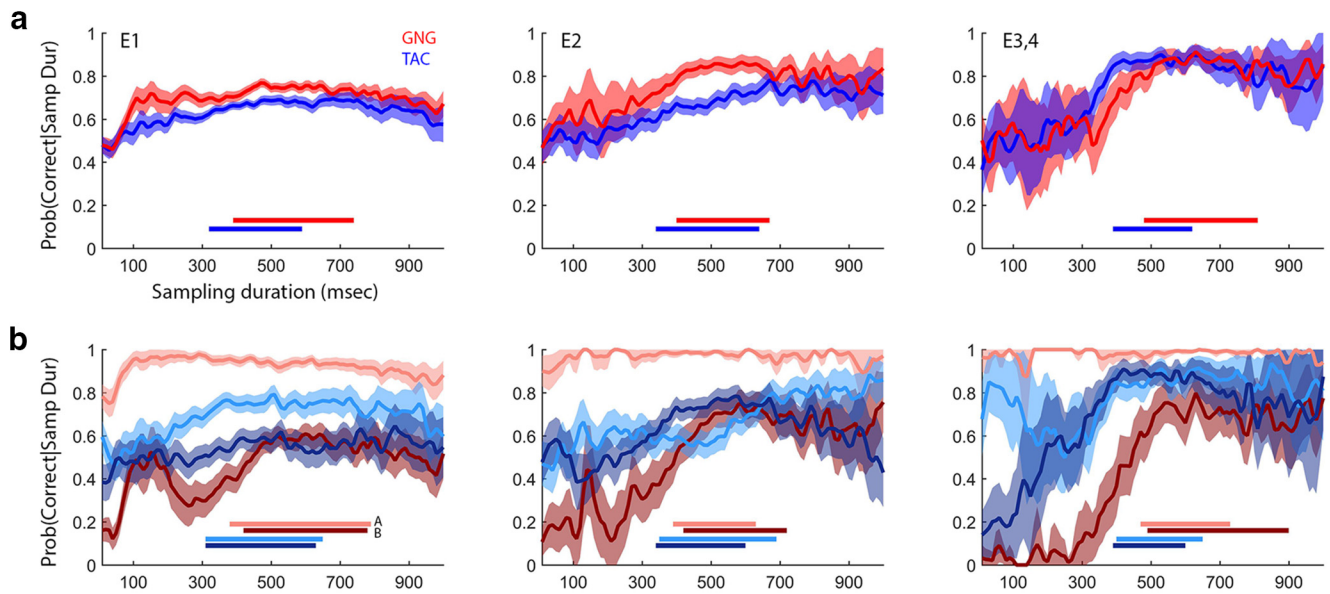
Figure 3. Performance conditioned on sampling duration. *a*, Empirical conditional probability plots for E1–E4 showing the probability of a trial being correct given a sampling duration, pooled by task (GNG, red; TAC, blue). Right, Data from E3 and E4 combined. Data are pooled within task across subjects and odor sets to show the mean space. Shaded area represents 95% binomial CIs on each sampling duration. Bars correspond to median (25th–75th percentile) sampling durations. *b*, Data pooled by odor (A or B) within each task. Lighter color, Odor A; darker color, odor B. Plots of the conditional probability were smoothed using MATLAB's smooth function with the lowess parameter.

between sampling duration and performance?). Although the experiments were originally designed to answer the first question, we have also attempted to offer some insight into the second question by looking at the relationship between sampling duration and performance.

Pooled empirical conditional probability plots (see Materials and Methods for discussion of how plots were made) showed an association between sampling duration and performance at the trial level that appears sigmoidal (Fig. 3), as expected from our previous research (Frederick et al., 2011). This association is consistent with past work in which mice were forced to sample longer than their normal sampling times (Rinberg et al., 2006) and supports the idea that sampling longer results in better performance at the trial level.

Rats tested in GNG show a ceiling-level performance for odor A (the go odor; Fig. 3b, pink curves). They achieve perfect or near-perfect performance for very short sampling durations. This is consistent with the bias that we reported above (see performance results). It appears that when tested in GNG, rats have a default go response such that when there is little or no information (as inferred from short sampling durations) the rats go. Rats tested in TAC do not appear to have the same level of default response bias, which is also consistent with the bias results reported above for TAC (see performance results).

We concluded that traditional logistic regression at the trial level would not adequately capture the underlying psychometric functions, which may be present within a session, because the independent variable (sampling duration) was not evenly sampled by each rat. This is supported by rats' narrow interquartile ranges in session-sampling duration [mean (SD) of 75th quantile–25th quantile; E1: 227 (121 ms), E2: 226 (85 ms), E3: 242 (132 ms), E4: 242 (95 ms)]. We are in the process of modeling this behavior, which is the focus of current and future work.

Discussion

Variability in past reports regarding olfactory operant decision making in rats argued both for and against accumulation of in-

formation over time (temporal integration) and changes in sampling time dependent on odor-discrimination difficulty. We focused on possible differences between GNG and TAC behaviors that had previously suggested different underlying problem-solving strategies (Uchida and Mainen, 2003; Abraham et al., 2004; Kay et al., 2006; Rinberg et al., 2006). Drawing from past work spanning behavior and electrophysiology, we hypothesized at least two ways in which rodents performed these tasks. We predicted that the low-reward, asymmetric response and reward structure in GNG would produce deliberative decision making and that rats would increase sampling to accumulate more information to perform better in harder discriminations. In contrast, due to the symmetric response and reward structure in TAC, we predicted rats would prefer speed to accuracy and sample odors less than in GNG, resulting in worse performance. We found that GNG rats sniff longer and perform better than TAC rats, supporting a more deliberative approach for GNG. However, reward structure is not the complete story, because manipulating reward rates did not erase this difference in E1, and the GNG performance advantage only occurred in the single-task condition (E1, E2). Rats performing either task did better with longer sampling, arguing for temporal integration in odor sampling in both tasks.

Task effects on behavior

We found that task type does matter, as do task parameters, contrary to our previous evidence based only on initial learning, which suggested that GNG and TAC could be arbitrarily similar (Frederick et al., 2011).

Sampling duration presented the most robust task difference and was significant in three of four experiments and in the same direction in the fourth (Fig. 2c,d; Table 2). Average sampling durations across all experiments were 505 ms for GNG (after subtracting the estimated 60 ms odor delay) and 410 ms for TAC, which is ~1 additional sniff for GNG. We observed longer sampling and decision times than other studies (200–375 ms; Uchida and Mainen, 2003; Abraham et al., 2004; Rinberg et al., 2006). One reason for this may be that our methods do not cap how long

rats can sample, and rats often sample much longer than the 1–2 s permitted in other studies (Fig. 2a). Our within-subjects tests (E3, E4) show that rats trained and tested on both tasks also sampled longer in GNG, even within the same session (E4). One reason for the lack of statistical difference in E2 could be because it was a between-subjects design with a smaller number per group than E1. Rats were also tethered for recording during the task, and this may have affected their natural movements.

It is reasonable to expect that sampling differences might be explained by sampling bias in which GNG rats sample no-go or go odors longer, but these effects were not consistent across experiments. In E2 and E3 rats sampled odor B longer (odor A − B; E2 = [−94, −45 ms], E3 = [−85, −22 ms]). However, in E1 they sampled A longer ([8, 49 ms]) and E4 showed no difference ([−45, 37 ms]). Therefore, it is possible that rats use an extra sniff to ensure a reward (CS+) or mobilize response inhibition (CS−) in some GNG contexts, but such strategies are not consistent across conditions. One alternative interpretation is that the background of response inhibition produces a cognitive load and slows decisions in GNG (Chambers et al., 2009).

Performance differences across tasks depended on experimental context. When rats were trained and tested on one task (E1 and E2), GNG rats performed better than TAC rats (Fig. 2b,d; Table 2). However, when rats learned both tasks (E3 and E4), subjects performed both tasks well (Fig. 2b). Rats performed TAC at GNG levels in E3 and E4 without sniffing as long as GNG (Tables 2, 3). Therefore, longer sampling durations may not be necessary for high performance in TAC when rats also know GNG. The GNG context could engage extra learning, which helps when performing TAC.

Rats altered sampling durations across odor sets in E1 (Table 2), as shown in previous studies (Abraham et al., 2004, 2012). We may have lacked power to detect differences in E2–E4 (E1 was our largest sample). Alternatively, it is possible that we did not use a large enough spread of discrimination difficulties to see this effect in all experiments (Extreme odor set was used only in E1 and E2). In E1 there was a significant difference between the shortest and longest sampling duration odor sets (S2 and Extreme), suggesting that rats adjust sampling times for discrimination difficulty even in TAC. That we did not see this difference in E2–E4 suggests that rats can do most monomolecular and mixture discriminations with similar sampling durations and extend times only when odors are very similar (Extreme). Another interpretation is that rats are satisfied with their performance for a given odor set once they reach some arbitrary performance level, what we have referred to as "good enough" performance (Kay et al., 2006), and this level may change dependent on task parameters and a trade-off with cognitive load.

Sampling duration was significantly correlated with performance at the session level in three of the four experiments, which argues for temporal integration in odor sampling. The one experiment (E1) in which we found no effect using GLS statistical models was that in which we tested rats for only 3 d on each odor set. We did not test these rats as long on each odor set as in E2–E4, so that we could probe performance and other variables given the same amount of training across tasks and odor sets. Temporal integration may show up more strongly in GLS statistics when rats perform more sessions and have a chance to increase performance more. However conditional probability plots for pooled trials in E1 do show evidence that performance increases with sampling duration (Fig. 3a.i,b.i).

The term temporal integration assumes that some type of information is accumulating over time. We have not proved

that this happens, although we have argued for it based on conditional probability plots (Fig. 3). It is indeed possible that sampling longer allows rats to make better decisions via some other mechanism.

Role of reward schedule
Reward schedule (half or full) had, by design, a large influence on reward frequency, but task and performance effects and other variables often erased predicted frequency differences. Response delays showed a possible interaction with apparently opposite effects on reward in the two tasks, but pairwise differences were not significant (Table 3). Reward schedule affected the time to initiate trials for GNG but not TAC; GNG rats initiated trials more slowly in the Full versus Half reward conditions (Table 4). Reward intervals also varied across tasks and reward conditions. TAC rats showed almost double the reward interval for Half relative to Full reward, as predicted from the experimental design, but GNG rats' reward intervals did not vary across reward conditions.

Thus, reward schedules within commonly designed GNG and TAC tasks can have complex effects on rats' behavioral strategies. That reward schedule did not influence performance or sampling duration was surprising given strong task differences in these measures and reward schedule effects on response and start delays, but such effects may emerge with different task parameters.

Different strategies, same cognitive system?
Conditional probability plots (Fig. 3) show what appear to be underlying psychometric functions for both tasks. Because rats did not choose sampling times uniformly across all durations, we cannot use these plots to show that for each subject sniffing longer results in better performance. We rely on pooled data for a more qualitative picture, and group-level GLS statistics provide evidence that rats do integrate olfactory information over time in both tasks. GNG and TAC strategies may then be viewed as different parameter ranges of the same underlying cognitive system, which may be supported by a unified underlying neural system (Frederick et al., 2016).

Our data show that many features of a task can affect behavioral outcomes, so we cannot claim that the results presented here are the final word on task differences in odor-discrimination behavior. It is possible that more extreme manipulations of reward probability, intertrial interval, and penalty delays could press rats to adopt different sampling strategies depending upon what, if anything, they are attempting to optimize. These tests are the object of future studies.

Importance of task selection
Our results support temporal integration across olfactory tasks, where sniffing longer gives rats a better chance of making a correct response. This shows that olfactory decisions are similar to other sensory decisions (Gold and Shadlen, 2002). Despite many similarities, there were important differences between GNG and TAC tasks, sampling duration being the most robust. These results help reconcile previous studies, performed in different laboratories, in which longer sampling times were seen in GNG tasks. That previous TAC studies did not find evidence of temporal integration suggests that an aspect of training or test parameters did not encourage subjects to use their entire behavioral range. These results show that more experiments and research should focus on the impact or interaction of the particular behavioral task on cognitive interpretations in research. Indeed, if we had only used one of the tasks, we would have come to radically

different conclusions about rats' behavioral strategies and abilities when performing odor discriminations. Researchers must take care to use multiple behaviors and parameters if they wish to draw general conclusions beyond a particular task.

References

Abraham NM, Spors H, Carleton A, Margrie TW, Kuner T, Schaefer AT (2004) Maintaining accuracy at the expense of speed: stimulus similarity defines odor discrimination time in mice. Neuron 44:865–876. CrossRef Medline

Abraham NM, Guerin D, Bhaukaurally K, Carleton A (2012) Similar odor discrimination behavior in head-restrained and freely moving mice. PLoS One 7:e51789. CrossRef Medline

Chambers CD, Garavan H, Bellgrove MA (2009) Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. Neurosci Biobehav Rev 33:631–646. CrossRef Medline

Clifford D, McCullagh P (2006) The regress function. R News 6:6–10.

Doucette W, Restrepo D (2008) Profound context-dependent plasticity of mitral cell responses in olfactory bulb. PLoS Biol 6:e258. CrossRef Medline

Frederick DE, Rojas-Líbano D, Scott M, Kay LM (2011) Rat behavior in go/no-go and two-alternative choice odor discrimination: differences and similarities. Behav Neurosci 125:588–603. CrossRef Medline

Frederick DE, Brown A, Brim E, Mehta N, Vujovic M, Kay LM (2016) Gamma and beta oscillations define a sequence of neurocognitive modes present in odor processing. J Neurosci 36:7750–7767. CrossRef Medline

Friedrich RW (2006) Mechanisms of odor discrimination: neurophysiological and behavioral approaches. Trends Neurosci 29:40–47. CrossRef Medline

Friedrich RW, Laurent G (2004) Dynamics of olfactory bulb input and output activity during odor stimulation in zebrafish. J Neurophysiol 91:2658–2669. CrossRef Medline

Gamble KR, Smith DW (2009) Discrimination of "odorless" mineral oils alone and as diluents by behaviorally trained mice. Chem Senses 34:559–563. CrossRef Medline

Gold JI, Shadlen MN (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 36:299–308. CrossRef Medline

Kay LM, Beshel J, Martin C (2006) When good enough is best. Neuron 51:277–278. CrossRef Medline

Laurent G (1996) Dynamical representation of odors by oscillating and evolving neural assemblies. Trends Neurosci 19:489–496. CrossRef Medline

McCullagh P, Clifford D (2006) Evidence for conformal invariance of crop yields. Proc R Soc A Math Phys Eng Sci 462:2119–2143. CrossRef

Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. New York: Springer.

Pinheiro J, Bates D, DebRoy S, Sarkar D (2013) nlme: Linear and nonlinear mixed effects models. proglangR package version 3.1–109.

Rinberg D, Koulakov A, Gelperin A (2006) Speed-accuracy tradeoff in olfaction. Neuron 51:351–358. CrossRef Medline

Rojas-Líbano D, Kay LM (2012) Interplay between sniffing and odorant sorptive properties in the rat. J Neurosci 32:15577–15589. CrossRef Medline

Schaefer AT, Margrie TW (2007) Spatiotemporal representations in the olfactory system. Trends Neurosci 30:92–100. CrossRef Medline

Skarda CA, Freeman WJ (1987) How brains make chaos in order to make sense of the world. Behav Brain Sci 10:161. CrossRef

Stanislaw H, Todorov N (1999) Calculation of signal detection theory measures. Behav Res Methods Instrum Comput 31:137–149. CrossRef Medline

Szucs D, Ioannidis JP (2017) Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLoS Biol 15:e2000797. CrossRef Medline

Uchida N, Mainen ZF (2003) Speed and accuracy of olfactory discrimination in the rat. Nat Neurosci 6:1224–1229. CrossRef Medline

Zariwala HA, Kepecs A, Uchida N, Hirokawa J, Mainen ZF (2013) The limits of deliberation in a perceptual decision task. Neuron 78:339–351. CrossRef Medline