Fentanyl reinforcement history has sex-specific effects on multi-step decision-making

Authors
Eric Garr[1,#], Yifeng Cheng[1,2,#,*], Laia Castell[1], Andy Dong[1], Cecelia Shuai[1], Yunru Wang[1], Jennifer Min[1], Zachary Camhi[1], Albert Nay[1], Nora Malloy[1], Patricia H. Janak[1,2,3*]

Affiliations
[1]Department of Psychological & Brain Sciences, Krieger School of Arts & Sciences, Johns Hopkins University, Baltimore, MD, 21218, USA
[2]Kavli Neuroscience Discovery Institute, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA
[3]Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, 21205, USA

[#]These authors contributed equally to this manuscript

*Corresponding authors (ycheng62@jhu.edu, patricia.janak@jhu.edu)

ABSTRACT

It is commonly thought that drug addiction involves a transition to habitual control of action, where the choice to consume drugs becomes automatized and reflects a failure to deliberate over possible negative outcomes. Determining whether the pursuit of addictive drugs is habitual is hampered by a lack of behavior assessments suitable for use during a bout of drug seeking. Therefore, to understand how variable histories of drug reinforcement might affect goal-directed and habitual pursuit of drug rewards, we trained rats to perform a multi-step decision-making task to earn oral fentanyl or sucrose rewards following extensive training with either fentanyl or sucrose. Importantly, this task allowed for independent measurements of goal-directed and habitual choice characteristics during online pursuit of rewards, and habitual choice could further be categorized into perseverative and reward-guided components. Extensive fentanyl training led to a bias for reward-guided habitual choice specifically in females, and a relatively high degree of perseveration in both sexes. These behavioral changes after chronic fentanyl generalized across fentanyl and sucrose seeking. In contrast, during fentanyl sessions, females showed increased perseveration, and both sexes showed blunting of the gradual improvement in goal-directed choice. These results show that chronic drug reinforcement induced habits that generalized across seeking drug and non-drug rewards, and female rats were especially susceptible to habitual control induced by both chronic and acute drug reinforcement.

INTRODUCTION

The drug-seeking behaviors of individuals with substance use disorders are colloquially referred to as "habits". The term "habit" has developed a specific definition which means an action that is insensitive to its future consequences (Dickinson, 1985). This definition has been applied in drug addiction research, as exposure to addictive drugs increases the likelihood of habitually seeking rewards (Corbit, Nie, & Janak, 2012; Corbit, Chieng, & Balleine, 2014; Nelson & Killcross, 2006; Renteria et al., 2021; Voon et al., 2015; Sebold et al., 2014). However, this area of research suffers from several major limitations. One is that the effect of drugs on habit formation is commonly measured in the context of seeking nondrug rewards (Corbit et al., 2014; LeBlanc, Maidment, & Ostlund, 2013; Nelson & Killcross, 2006; Nordquist et al., 2007; Renteria et al., 2021). This precludes the study of a drug habit per se. There have been attempts to overcome this limitation by delivering the drug as the instrumental reward on second-order schedules and then attempting to elicit a habit through overtraining (Belin & Everitt, 2008; Hynes et al., 2024; Murray et al., 2015), but this procedure makes it difficult to infer whether responding for drug is outcome-insensitive. Another approach has been to deliver the drug as the instrumental reward and then assess habitual responding by subsequently devaluing the drug reward (Corbit, Nie, & Janak, 2012; Jones et al., 2022) or devaluing the instrumental contingency by extinguishing the drug-taking response (Giuliano et al., 2021; Zapata, Minney, & Shippenberg, 2010), and then testing instrumental performance in extinction. However, the devaluation approach also has limitations. For example, an action that is insensitive to devaluation can create ambiguous interpretations. Such an action could be a manifestation of either a "model-free" (MF) mechanism, where actions are driven solely by an expectation of past rewarding outcomes, or perseveration, where action repetition is not based on outcome expectations of any kind (Miller, Shenhav, & Ludvig, 2019). These two types of inflexible behavior, possibly orthogonal, cannot be distinguished by devaluation methods. Additionally, the use of reward devaluation makes it difficult to know whether drugs promote habit or whether they impair goal-directed control (Vandaele & Janak, 2018), and this limitation hinders the translational validity of the habit model of drug seeking because it calls into question whether drug-seeking behaviors are habitual (Vandaele & Ahmed, 2021). Therefore, a different experimental approach is needed.

The two-step task is an alternative but commonly used method for eliciting goal-directed and habitual components of action control (Daw et al., 2011), and lends itself to a suite of analyses that can independently measure these aspects of performance during online reward seeking. One strategy that is sometimes used to solve the two-step task, referred to as model-based (MB) learning, requires an agent to represent the probabilities between actions and outcomes such that, when faced with a choice between actions, the rat is able to plan and find the optimal route connecting action with future reward. In contrast, an MF strategy requires an agent only to represent the long-run probabilities of reward associated with each action without accounting for the route connecting action to reward. MB and MF decision-making are often identified with goal-directed and habitual control of action, respectively (Dolan & Dayan, 2013). However, MF choice can vary independently of perseveration in the two-step task, which makes for an additional source of habitual behavior. While MF control of choice is sensitive to prior rewards but insensitive to future action transitions, perseveration is insensitive to both. These features make the two-step task well-suited for addressing how online drug reinforcement alters goal-directed, habitual, and perseverative components of decision-making. The task is most commonly used in human

participants but has been expanded to mice (Akam et al., 2021; Blanco-Pozo, Akam, & Walton, 2024) and rats (Groman et al., 2019; Hasz & Redish, 2018; Miller, Botnivick, & Brody, 2017).

Here, we asked how variable histories of fentanyl reinforcement affect MB, MF, and perseverative aspects of choice behavior on a two-step task in rats. Fentanyl is a potent synthetic opioid that plays a significant role in the current opioid epidemic. Previous research has shown that chronic exposure to opioids, including fentanyl, disrupts cognitive flexibility (Halbout et al., 2024; LaRocco et al., 2024; Ma et al., 2019; Tamargo et al., 2021; Wakhlu et al., 2023; but see Wheeler et al., 2023), but it is less clear whether and how acute fentanyl exposure changes cognitive flexibility during ongoing decision-making, and what decision processes underly fentanyl-seeking. In addition, although there is evidence that opioids bias behavior toward perseveration (Boulos et al., 2019; Hölter, Tzschentke, & Schmidt, 1996; Korin, 1974; Loh, Smith, & Roberts, 1993; Lyvers & Yakimoff, 2003; Sanchez-Roige, Ripley, & Stephens, 2015; Seip-Cammack & Shapiro, 2014), it is unknown whether an action reinforced with opioid reward would show MF features in addition to well-known perseverative features, and whether the conditions that support opioid-induced perseveration do so in a way that is dissociated from MF choice.

We therefore trained rats over a prolonged period of days to perform a two-step task to earn oral fentanyl or sucrose rewards, followed by a brief period of fentanyl and sucrose sessions in alternation. This experimental design allowed us to dissociate the effects of chronic and acute fentanyl reinforcement on different components of learning and decision-making. Moreover, because biological sex influences how opioids impact cognitive function, with females being more vulnerable to opioid-induced impairments in cognitive flexibility (Anderson et al., 2021; LaRocco et al., 2024; Wakhlu et al., 2023), we hypothesized that the effect of chronic and acute fentanyl would be greater in females than males.

Here, we show that chronic and acute fentanyl reinforcement cause distinct disruptions in flexible multi-step decision-making. Chronic fentanyl induced a high degree of perseveration in males and females that generalized across fentanyl and sucrose seeking, and also induced an MF bias specifically in female rats that also generalized across fentanyl and sucrose seeking. On the other hand, acute fentanyl induced changes in decision-making that were specific to fentanyl and did not generalize to sucrose—acute fentanyl interfered with a progressive improvement in MB choice and also increased perseveration in females given brief, but not extensive, fentanyl training.

RESULTS

Rats show model-based control of action on a two-step task

To determine the task design that most effectively yields goal-directed control of action that can be compared against subsequent performance under fentanyl reinforcement, we first compared rats' performance on two different versions of the two-step task for sucrose rewards. Rats initiated trials by making a nose-poke in a center magazine, after which they could express their choice by nose-poking into either the left or right port (**Figure 1A**). Once the nose-poke was detected, one of two levers inserted in the opposite side of the chamber and was accompanied by either a high or low

pitch tone. After the rat pressed the extended lever, the lever retracted, and the trial outcome was delivered: liquid sucrose (accompanied by a click) or reward omission (accompanied by white noise). Each lever was associated with different reward probabilities (0.8 and 0.2), which were periodically reversed throughout each session. The two different versions of the task—hereafter referred to as "simple" and "complex"—were distinguished by how the transition probabilities were controlled (**Figure 1B**). Specifically, in the simple task, the transition probabilities that governed which nose-poke led to which lever insertion were held constant, while in the complex task, the transition probabilities reversed probabilistically. Both tasks have been used in prior rodent studies (Akam et al., 2021; Blanco-Pozo et al., 2024; Groman et al., 2019; Hasz & Redish, 2018; Miller et al., 2017) but have never been directly compared.

Thirsty rats were trained on both tasks in a balanced order. In response to reversals in reward probabilities and transition probabilities, rats adjusted their choices accordingly (**Figures 1C and 1D**). For reward probability reversals, rats adjusted their choices following reversals in both tasks (**Figure 2A and 2B**; main effect of reversal phase, $F(1,17) = 141.02$, $p < 0.001$), although the simple task encouraged more optimal behavior by increasing the probability of choosing the nose-poke that led to a higher probability of reward (task x reversal phase interaction, $F(1,18) = 29.52$, $p < 0.001$). Notably, when rats performed the simple task, they performed fewer trials and achieved a higher reward rate (**Figure S2A and S2B**). Across both tasks, rats showed similar overall probabilities of left and right choices (**Figure S2C**) and latencies to initiate the first and last trials (**Figures S2D and S2E**), which indicate similar action bias and motivation levels across tasks.

To differentiate between MB and MF action strategies, we conducted an analysis on the rats' stay probabilities—defined as the probability of repeating the same nose-poke choice as the previous trial as a function of the previous trial's transition and outcome. Each strategy leads to a unique pattern of conditional stay probabilities (**Figure 2D**). In short, MB choice considers the interaction between the type of transition experienced (common or rare) and the outcome (reward or omission), while MF choice only considers the outcome. Rats exhibited distinct patterns of stay probabilities that differed between the two tasks (**Figure 2F**; task x transition x outcome interaction, $F(3,51) = 12.11$, $p < 0.001$). The pattern of stay probabilities in the complex task was consistent with MF choice—a higher probability of repeating choices that were previously rewarded regardless of the transition type. Conversely, stay probabilities in the simple task were consistent with a mix of MB and MF choice. An MB index was derived by comparing the stay probabilities after common transitions with rewards and rare transitions without rewards against the opposite combinations, and an MF index was calculated by comparing the probability of staying after any rewarded trial against any unrewarded trial (see *Methods*). A comparison of MB and MF indices showed that rats exhibited a significantly higher MB index in the simple task compared to the complex task (**Figure 2G**; $t(18) = 4.85$, $p < .001$), while the MF index was significantly lower in the simple task (**Figure 2G**, $t(18) = 3.04$, $p = 0.007$). Additional analyses that regressed choice against trial-level variables multiple trials back in time also showed superior MB choice during performance of the simple task (**Figure S2**).

Overall, rats exhibited enhanced MB choice and diminished MF choice during performance of the simple two-step task. We hypothesized that extensive fentanyl reinforcement would decrease and increase MB and MF choice, respectively, and therefore, the pattern of choices engendered by the

simple task served as a useful baseline against which to compare performance under fentanyl reinforcement.

Oral fentanyl spares reversal learning

To understand how distinct fentanyl reinforcement histories affect decision-making in the two-step task, we trained a separate cohort of thirsty rats to earn either oral fentanyl or sucrose rewards on the simple two-step task (i.e. without transition reversals). The training period lasted 22-30 sessions, which ensured that rats fully acquired knowledge about the task structure. Training was then followed by a brief period in which rats could earn fentanyl or sucrose in alternating sessions (**Figure 3A**). This period lasted 12 sessions (i.e. 6 sessions with each reward type). This experimental design allowed us to investigate how variable drug reinforcement histories impact decision-making in response to drug and natural rewards. All analyses were conducted on the final six sessions of alternating rewards (i.e. 3 sessions with each reward type), unless otherwise noted.

We first confirmed that oral fentanyl is reinforcing beyond its thirst-quenching properties by training rats to press a lever for either oral fentanyl or water during conditions of thirst and non-thirst (**Figure 3B**). Rats escalated their intake of fentanyl more quickly than water regardless of deprivation state (main effect of reward: $F(1,7) = 9.80$, $p = .017$; no reward x deprivation interaction: $F(1,7) = .031$, $p = .866$). We additionally confirmed that instrumental responding for oral fentanyl depends on the activation of opioid receptors (**Figure 3C**). Thirsty rats were trained to press a lever for either oral fentanyl or sucrose after an injection of the opioid antagonist naltrexone, or saline. Naltrexone significantly affected pressing for fentanyl, but not sucrose (reward x injection interaction: $F(1,7) = 7.05$, $p = .033$). Specifically, rats took longer to terminate pressing for fentanyl after naltrexone relative to saline, most likely because the drug intake threshold was higher (**Figure 3C**). These results show that oral fentanyl motivates instrumental responding in a way that depends on opioid receptor activation.

As a first step, we sought to understand whether fentanyl alters a simple form of flexible decision-making during performance of the two-step task by analyzing reversal learning. This analysis cannot provide direct insight into MF or MB strategies, but can indicate the extent to which rats incorporate reward probabilities into their choices. We analyzed reversal learning by plotting nose-poke choice as a function of the relative time from reward probability reversal—that is, when the reward probabilities associated with each lever abruptly switched (**Figures 3D and 3E**)—and then averaging over trials 1-5 before ('pre' block) reversal and trials 15-20 after ('post' block). A reward type x block x training group x sex ANOVA revealed a main effect of block ($F(1,31) = 63.14$, $p < .001$), but block did not interact with reward type, training group, or sex (F's$(1,31) < 1.79$, p's $> .191$). This analysis shows that reward probability reversals caused rats to switch their nose-poke preferences, and the lack of any interactions with group or instrumental reward type implies that nose-poke choices were sensitive to future reward probabilities to a similar degree whether fentanyl or sucrose served as the reward, and whether fentanyl reinforcement history was extensive or brief.

In females, extensive training with fentanyl biases behavior towards model-free choice

To examine how fentanyl reinforcement history affects the balance between MF and MB choice, we first analyzed the session-averaged stay probabilities (**Figure 4A**). A mixed-model ANOVA was conducted with outcome (reward vs. omission), transition (common vs. rare), instrumental reward (fentanyl vs. sucrose), training group (fentanyl vs. sucrose), and sex (male vs. female) as factors. This analysis revealed a main effect of outcome ($F(1,31) = 55.28$, $p < .001$), an outcome x transition interaction ($F(1,31) = 14.70$, $p < .001$), and an outcome x transition x group x sex interaction ($F(1,31) = 9.21$, $p = .005$). The main effect of outcome arose from a greater stay probability following reward vs. omission trials. This is consistent with MF choice since MF choice is, by definition, guided only by long-run probabilities of reward without regard for the transitions connecting action and reward. However, simple main effects tests showed that the source of the outcome x transition interaction was an MB pattern of stay probabilities (**see Figure 2D**): a greater stay probability after reward-common vs. reward-rare trials ($p = .01$), and a greater stay probability after omission-rare vs. omission-common trials ($p = .006$). These results show that, overall, rats' choices were a mixture of MF and MB strategies, consistent with what we found earlier when comparing the simple and complex tasks (**see Figures 2F and 2G**).

To understand the source of the four-way interaction between outcome, transition, group, and sex, we computed MF and MB indices for all subjects and analyzed as a function of reward type, training group, and sex. This analysis yielded a significant index x group x sex interaction ($F(1,31) = 6.13$, $p = .019$), and simple main effects tests showed a significantly higher MF index in female rats given prior training with fentanyl ($p = .009$; **Figure 4A**). We applied the same analyses to multi-trial regression coefficients and found similar results (**Figure S3**). These analyses show that, overall, there was a mixture of MF and MB choice across rats, with the balance biased in favor of MF for female rats given extensive training with fentanyl.

Female rats are more susceptible to the acute effects of fentanyl on perseveration

In addition to the interactions with trial-level variables reported above, the analysis of stay probabilities also showed a main effect of group ($F(1,31) =4.75$, $p = .037$), a reward x group interaction ($F(1,31) = 5.98$, $p = .02$), and a reward x sex interaction ($F(1,31) = 4.78$, $p = .037$). These statistical results imply that stay probabilities, regardless of trial-level variables like outcome and transition, depended on the history of fentanyl reinforcement, the acute online presence of fentanyl, and sex. When stay probabilities are averaged over trial-level variables, they can be thought of as measuring perseveration—the overall tendency to repeat actions. We did not find any significant correlations between perseveration and MF or MB indices (**Figure S4)**, implying that perseveration is a separate dimension of choice.

To understand the source of these significant interactions, the average stay probabilities were plotted as a function of group, reward, and sex (**Figure 4B**). Perseveration was greater for the group trained with fentanyl than sucrose (**Figure 4B**), but acute fentanyl reinforcement increased perseveration even for the group trained with sucrose (**Figure 4B**). In addition, among rats that received prior extensive training with sucrose, females perseverated more during fentanyl than sucrose sessions, while this was not true of males (**Figure 4B**). Putting these results together, extensive training with fentanyl increased perseveration in a way that generalized across reward identity and sex, while the perseveration observed within fentanyl-reinforced sessions after chronic sucrose was highly specific to reward identity and sex. Specifically, females showed

greater choice perseveration during acute fentanyl in a way that did not generalize to sucrose (**Figure 4B**). Interestingly, there were no significant correlations between perseveration and MF bias in females (**Figure 4C**; $-.46 < r$'s $< .12$, $.25 < p$'s $< .99$), which supports the idea that extensive and brief fentanyl reinforcement histories have distinct behavioral effects in female rats.

It could be argued that these analyses do not precisely assess the acute effects of fentanyl in isolation from their chronic effects because we ignored data from the first few sessions when rats were first exposed to fentanyl (see *Methods*). The reason we chose to ignore these sessions is because rats were becoming familiar with the new rewards and altered contexts, which could have interfered with performance of the task. Nevertheless, we chose to examine the very first fentanyl and sucrose sessions for rats given prior training with sucrose to uncover whether acute fentanyl altered perseveration relative to sucrose during the first exposure to the drug. Indeed, perseveration was greater during the first fentanyl session compared to the subsequent sucrose session, although there was no interaction with sex (**Figure S5**). On average, perseveration increased over the course of the fentanyl session and decreased during the sucrose session (**Figure S5**), but this difference was not evident in a reward x session half interaction ($F(1,16) = 2.28$, $p = .150$). However, the degree of perseveration during fentanyl sessions was positively correlated with the number of fentanyl rewards earned ($r = .47$, $p = .048$). This was not true of sucrose ($r = .13$, $p = .62$). This analysis shows that the effect of fentanyl reward on perseveration was evident from the first encounter with fentanyl, and that this is likely a pharmacological effect.

Acute fentanyl reinforcement stifles within-session improvement in model-based choice

When rats earned fentanyl rewards on the two-step task, they accumulated fentanyl over the course of a session, and therefore, session-averaged behavioral measures may not provide a full description of how fentanyl affects choice. In addition to session-averaged stay probabilities, we analyzed within-session changes in stay probabilities by breaking down the session by halves (**Figure 5A**). In addition to replicating the main effects and interactions reported above, a mixed model ANOVA that included session halves as a factor revealed an outcome x transition x reward x half interaction ($F(3,93) = 4.27$, $p = .047$). Simple main effects tests showed that, during the second half of the session, stay probabilities following reward-common trials were smaller during fentanyl versus sucrose sessions ($p = .006$), while stay probabilities following reward-rare trials were greater during fentanyl versus sucrose sessions ($p = .031$).

To visualize this complex effect, we computed reward-guided MB indices by subtracting the stay probabilities between reward-common trials and reward-rare trials. Across training groups and sexes, there was a clear increase in reward-guided MB choice over the course of sucrose sessions, while this behavior decreased during fentanyl sessions (**Figure 5B**). Notably, session halves were divided by trials, not time. This means that each session half should contain roughly equal numbers of rewards, which would mean that the time-dependent differences in MB choice for fentanyl versus sucrose are likely not due to differences in reward rate or motivation. Indeed, there was no significant difference in the number of rewards earned between the first and second session halves ($F(1,34) = 2.49$, $p = .124$). Perseveration did not show any dynamic changes across the session (**Figures 5C and 5D**; no main effect or interactions with session half, p's >

.143). Thus, a component of model-based choice for sucrose reward showed a progressive improvement within sessions—an effect that was not observed for fentanyl reward. This trend generalized across training groups and sexes, suggesting it is an acute effect, and not dependent on prior chronic fentanyl.

DISCUSSION

We find that fentanyl-seeking actions are driven by different combinations of model-based, model-free, and perseverative features depending on the history of fentanyl reinforcement and biological sex. Choices that led to fentanyl were largely model-free in female rats, but only when they were given extensive prior training with fentanyl. However, this type of habitual behavior also extended to sucrose seeking, which suggests that this is not a "drug habit" per se but rather a general change in decision-making. The finding that most resembled a drug-selective habit was the within-session decrement in model-based choice induced by cumulative fentanyl intake across sexes and training groups. Yet given the alignment between model-based choice and goal-directed action selection, it would be more accurate to state that acute fentanyl impaired goal-directed control rather than inducing a change in habits. We also found that model-free choice was largely independent of perseveration, both in terms of how the two variables map onto orthogonal dimensions and the drug reinforcement conditions that give rise to their changes. Female rats given brief training with fentanyl showed an acute increase in perseveration during fentanyl relative to sucrose sessions, and this change in perseveration was not accompanied by any changes in model-free choice. Furthermore, the female-specific increase in perseveration during fentanyl reinforcement was present during the very first session with fentanyl and correlated with the number of fentanyl rewards, which implies a pharmacological effect. This work demonstrates that drug-induced habits can arise from either a model-free or perseverative mechanism, with chronic fentanyl giving rise to perseveration across sexes and an additional increase in model-free choice specifically in females.

We first showed that training rats on a two-step task without transition reversals for sucrose rewards yields higher and lower degrees of model-based and model-free control of action, respectively, compared to a two-step task that contains transition reversals. This aligns with previous studies, which suggest that tasks without transition reversals favor model-based strategies in rodents (Groman et al., 2019; Miller et al., 2017). The simpler task likely allows rats to develop a more consistent internal model of the task, facilitating their ability to predict outcomes and optimize choices based on prior experience. In contrast, the complex task, which introduced reversals in both transition and reward probabilities, appears to have overwhelmed the rats' capacity to rely on model-based planning.

There are two possible explanations that explain the difference in decision-making strategies between the different two-step tasks. First, the high volatility of the complex task might have prevented the rats from forming a stable internal representation of the task, thereby impairing model-based learning. It is known that excessive uncertainty and complexity can cause a reversion to a model-free strategy (Kim et al., 2019). Second, it has been argued that model-based behavior can sometimes be generated by sophisticated model-free processes, which transition reversals can penalize (Akam, Costa, & Dayan, 2015). It is therefore possible that the model-based behavior that we observed in our simpler task is a byproduct of a complex model-

free learning process. Future research would benefit from further study with experimental designs that can differentiate between the two explanations of model-based behavior.

To understand how chronic and acute fentanyl reinforcement affect the balance between model-free and model-based choice, we next trained thirsty rats to earn oral fentanyl and sucrose rewards in separate sessions after extensive training on either fentanyl or sucrose. We chose to deliver fentanyl as an oral reward to compare with a control oral sucrose condition. While the most common preclinical model of fentanyl seeking is intravenous self-administration, oral delivery of fentanyl has been shown to reinforce an instrumental action in a way that mirrors the effects of intravenous fentanyl: inelastic demand curves, escalation of intake, and cued reinstatement (Coffey et al., 2023). As well, oral intake is a common route of administration for humans. Importantly, we show that the reinforcing properties of oral fentanyl in thirsty rats at least partially depend on the activation of opioid receptors, suggesting that reward-seeking is driven not only by the motivation to seek liquid while thirsty but also by the pharmacological effects of fentanyl itself.

Our experiment showed that decision-making processes in male and female rats are affected differently by chronic and acute fentanyl reinforcement. While chronic fentanyl reinforcement led to a high level of perseveration that generalized across sexes, the same drug history led to an increase in model-free choice for both fentanyl and sucrose only in female rats. There is some evidence in the literature that females respond differently to chronic fentanyl compared to males. After many repeated fentanyl exposures, female rodents self-administer fentanyl at higher rates, are more prone to fentanyl relapse, and show greater resistance to extinction of fentanyl seeking (Bakhti-Suroosh, Towers, & Lynch, 2021; Coffey et al., 2023; Hammerslag et al., 2021; Klein, Popke, & Grunberg, 1997; Little & Kosten, 2023; Malone et al., 2021; Towers, Setaro, & Lynch, 2022; but see Nicolas et al., 2022). The basis of these sex differences may lie in neuroendocrinology, as estrogen treatment leads to greater motivation and relapse for fentanyl in female rats (Towers, Setaro, & Lynch, 2023). The present study extends this work to show sex-specific effects of chronic fentanyl-induced changes in action control more broadly. Given that women generally become more rapidly addicted to opioids and are more susceptible to relapse (Goetz, Becker, & Mazure 2021), this impairment could be particularly relevant in understanding why females may be at greater risk for opioid dependence.

These findings of behavioral change after chronic fentanyl contrast with those observed after chronic sucrose within fentanyl-reinforced sessions. After extensive sucrose training, female rats showed an acute increase in perseveration during fentanyl-seeking relative to sucrose-seeking, despite displaying balanced model-free and model-based choice. This acute fentanyl-induced perseveration is probably a direct result of bioactive fentanyl, as perseveration was also high during the first session of fentanyl reinforcement in drug-naïve rats. Another manifestation of acute fentanyl was that, across sexes and in both chronic fentanyl and sucrose training groups, the accumulation of fentanyl within a session led to a decrease in model-based control. In contrast, model-based control gradually improved during sucrose seeking, likely due to improved task representation with accumulated experience. An acute disruption in behavioral flexibility may increase the risk of developing future drug abuse, as compromised cognitive flexibility is associated with greater vulnerability to addiction and predicts future escalation of drug use (Chen et al., 2021; Groman et al., 2020; LaRocco et al., 2024; Rhodes et al., 2020; Villiamma, Casby,

& Groman, 2022). It is uncertain whether compromised cognitive flexibility causes addictive behaviors or whether they are symptoms of the same underlying cause, although addressing this question would be a fruitful avenue for future research that seeks to understand the mechanisms underlying the transition from recreational drug use to addiction.

In summary, we found that rats can perform a multi-step decision-making task that yields measures of model-based, model-free, and perseverative choice characteristics that are modulated by chronic and acute fentanyl reinforcement. It would not be possible to unearth these behavioral differences with the standard methodology of measuring goal-directed and habitual control of action under free-operant schedules, highlighting the utility of combining the two-step task with drug reward for future directions in drug habit research. We found that females do indeed show greater evidence of model-free control of drug-seeking after chronic drug reinforcement, but this change in behavioral control was not limited to drug pursuit, but was observed across rewarding outcomes. We also found that accumulating within-session drug reduces goal-directed control during that session in both sexes and even in subjects that did not receive chronic drug. Relatedly, females that did not receive prior chronic fentanyl showed greater perseveration during fentanyl-reinforced sessions. Our results highlight a complex impact of fentanyl on behavioral control with dissociable effects following chronic and acute drug reinforcement. While acute drug exposure enhances perseveration in a sex-dependent manner, repeated drug reinforcement induces a habit that is largely perseverative across sexes, but also extends to model-free control in females. Overall, the finding that the length of prior fentanyl use and biological sex interact to produce distinct behavioral deficits may help to explain why some individuals transition from recreational drug use to addiction. Future research will benefit from understanding how distinct features of action control map onto and predict different features of addiction.

MATERIALS AND METHODS

Animals

Three different cohorts of rats were used. Twenty-four Long-Evans rats (Evingo; 18 male, 6 female) were used for the experiment that compared performance on the two types of two-step tasks. Thirty-five Long-Evans rats (Evingo; 19 males, 16 females) were used for the experiment that compared performance on the two-step task for fentanyl and sucrose rewards. Eight Long-Evans rats (Evingo; 4 males, 4 females) were used for the experiments that compared instrumental responding for fentanyl and water, and also fentanyl and sucrose after naltrexone injections. Rats were kept in a temperature- and humidity-controlled environment with a light:dark cycle of 12:12 h (lights on at 7:00 a.m.). Experiments were conducted during their light cycle. All animal care and experimental procedures were approved by the Johns Hopkins University Animal Care and Use Committee.

Two-step task

Rats were trained in a two-step decision-making task within a sound-attenuated operant chamber. The chamber was equipped with a reward magazine flanked by two levers on the south wall and a center magazine flanked by two nose-poke ports on the north wall. To motivate task performance,

rats were water-restricted throughout training. All rats underwent gradual shaping across four stages:

1. **Habituation:** On the first day, rats were allowed to freely explore the chamber and collect 30 rewards (0.1 ml of 10% sucrose solution) delivered randomly at an average interval of 60 seconds to the reward magazine on the south wall.
2. **Lever press training:** Rats were then trained to press the levers on the south wall under a fixed-ratio 1 (FR1) schedule until they earned a total of 100 rewards (0.05 ml per reward).
3. **Trial initiation training:** In the third phase, rats were trained to poke into the center magazine on the north wall, triggering the insertion of a single lever on the opposite wall. The rats could then press this lever to receive a reward.
4. **Simplified task**: At this stage, a simplified version of the two-step task was introduced. Each trial began with the illumination of the center magazine on the north wall. Upon entry into this magazine, one of the nose ports was illuminated, requiring a nose-poke as the first-step choice. This action triggered a distinct tone: a high pitch indicated the insertion of the left lever, while a low pitch indicated the right lever (second-step state). Each nose port was predominantly associated with one lever, with transition probabilities fixed at 0.8/0.2. After pressing the lever, the reward was delivered with 100% certainty, accompanied by a clicker sound. A fixed 3-second inter-trial interval (ITI) was maintained throughout this stage.

Final task (simple version): In the simple version of final two-step task, the structure of the task remained the same as in the last stage of shaping. Each trial began with the illumination of the center magazine on the north wall, prompting the rat to nose-poke as the first-step choice. This choice triggered the insertion of one of the two levers on the south wall, guided by the same transition probabilities (0.8/0.2) as in the previous stage. However, unlike the final stage of shaping where reward delivery was certain, the final task introduced probabilistic reward outcomes. When the rat pressed the lever, the reward was delivered with a probability of 0.8 if the rat had selected the high-reward lever, and 0.2 if the rat had selected the low-reward lever. A clicker sound accompanied reward delivery, while 1-s white noise was presented if reward was omitted. The reward probability associated with each lever was randomly alternated after a minimum of 20 trials, with a 10% chance of switching on each subsequent trial, and a maximum of 35 trials before a switch occurred. Reward size was fixed at 0.05 ml.

Final task (complex version): In the complex version of final two-step task, we introduced reversals in the transition matrix linking the first-step choice to the second-step state. These reversals occurred every 80-100 trials, with at least 20 trials separating transition reversals and reward probability reversals (**Figure 1C**).

Comparing performance on the two-step task with and without transition probability reversals

The order in which rats were trained on the simple and complex tasks was balanced. For one group of rats, training began with the simple version of the task, where they underwent 30-40 sessions to establish consistent task performance. After this period, these rats were transitioned to the complex version of the task, where they completed another 30-40 sessions. Conversely, a second group of rats was initially trained on the complex version of the task. These rats first completed 2-5 brief sessions of the simple task to familiarize them with the basic task structure before proceeding to

30-40 sessions of the complex task. After completing the complex task training, these rats were then trained again on the simple version of the task to assess how the experience with the complex task influenced their performance on the simpler task.

Free operant responding for fentanyl with and without fluid restriction

Long-Evans rats (4 males, 4 females) were water-restricted and trained to press two levers on a fixed-ratio 1 schedule for oral fentanyl (fentanyl hydrochloride from Cayman Chemical Company; 50 µg/ml, 0.05 ml per delivery) over two daily sessions. Then, each rat was given the opportunity to press one lever for fentanyl and the other lever for tap water while thirsty or not thirsty during separate sessions. The testing order of instrumental outcome and deprivation state were counterbalanced with sex.

Free operant responding for fentanyl after naltrexone injections

The same eight rats described above were then placed back on water restriction and trained to press one lever on a fixed-ratio 5 schedule for oral fentanyl (50 µg/ml, 0.05 ml per delivery) and another lever for sucrose (10%, 0.05 ml per delivery) over two daily sessions. Each rat was then given the opportunity to press each lever for their respective outcomes 30 minutes after IP injections of naltrexone (0.1 mg/kg) or saline. The testing order of instrumental outcome and injection type was counterbalanced with sex. Rats were given 1 hour of supplemental water after each session.

Comparison of performance on the two-step task during fentanyl reinforcement

A separate cohort of rats was trained to perform the two-step task to earn oral fentanyl (25 µg/ml; 10 males and 9 females) or 10% sucrose (10 males and 8 females). Rewards were delivered in 0.05 ml aliquots. The fentanyl concentration was lowered to 25 µg/ml to increase the number of trials the rats were willing to perform. Following shaping, training on the full task commenced and continued for 22-30 daily sessions. Rats were trained without transition probability reversals. During each session, the fentanyl group was free to earn up to 150 rewards, while the sucrose group was limited to the average number earned by the fentanyl group during the previous session. This was done to equate the average number of rewards between groups.

Following training, all rats received alternating sessions with fentanyl and sucrose rewards (6 each, 12 total). Sessions with the unfamiliar reward occurred in an altered context with lemon scent and honeycomb textured floors. For individual rats, the maximum number of sucrose rewards per session was set to the number of fentanyl rewards earned during the previous session. Only the final six sessions were analyzed to exclude the sessions during which rats became familiar with the novel reward and context not experienced during training.

Data analysis

Stay probability analysis

To assess decision-making strategies, we analyzed the probability of rats repeating their previous first-step choice ("stay probability") based on the outcome and transition of the previous trial. Trials were classified into four conditions: (1) common transition with reward, (2) common transition with reward omission, (3) rare transition with reward, and (4) rare transition with reward omission. Stay probability was calculated as the proportion of trials where the rat repeated the same first-step choice as in the previous trial.

To calculate the model-based (MB) and model-free (MF) indices, we used the following formulas:

<u>Model-based (MB) index:</u> This index was calculated as the sum of the probability of staying on the same choice following a reward-common trials (RC) and an omission-rare trials (UR), minus the sum of the probability of staying on the same choice following an omission-common trial (OC) and a reward-rare trial (RR). Mathematically, this is represented as:

$$MB = [p(stay|RC) \ + \ p(stay|OR)] - [p(stay|OC) \ + \ p(RR)]$$

<u>Model-free (MF) index:</u> This score was calculated as the sum of the probability of staying on the same choice following a reward, regardless of the transition type, minus the sum of the probability of staying on the same choice following no reward, regardless of the transition type. Mathematically, this is represented as:

$$MF = [p(stay|RC) \ + \ p(stay|RR)] - [p(stay|OC) \ + \ p(stay|OR)]$$

Multi-trial-back regression analysis

To assess the influence of past transitions and outcomes on future decisions, we conducted a logistic regression analysis that considers the effects of multiple previous trials (Miller et al., 2016; Miller et al., 2017). Similar to the last trial stay probability analysis, for each of the four possible trials (RC, RR, OC, OR), we created vectors representing these outcomes and transitions. A trial was coded as +1 if the left port was chosen, and -1 if the right port was chosen. The following regression model was employed:

$$\log\left(\frac{p_{left}(t)}{1 - p_{right}(t)}\right) = \sum_{a=1}^{T} \beta_{RC}(a) * RC(t-a) + \sum_{a=1}^{T} \beta_{RR}(a) * RR(t-a) + \sum_{a=1}^{T} \beta_{OC}(a) * OC(t-a) + \sum_{a=1}^{T} \beta_{OR}(a) * OR(t-a)$$

In this model, the regression coefficients $\beta_{RC}, \beta_{RR}, \beta_{OC}$, and β_{OR} capture the likelihood of repeating a choice made a trial ago, contingent on the specific transition and outcome type, while T represents the number of past trials considered by the model for predicting future choices. To compute MB and MF indices that account for influences from multiple past trials, we summed the relevant regression coefficients:

$$MB = \sum_{a=1}^{T}[\beta_{RC}(a) + \beta_{OR}(a)] - \sum_{a=1}^{T}[\beta_{OC}(a) + \beta_{RR}(a)]$$

$$MF = \sum_{a=1}^{T}[\beta_{RC}(a) + \beta_{RR}(a)] - \sum_{a=1}^{T}[\beta_{OC}(a) + \beta_{OR}(a)]$$

To assess the influence of past choices, correct actions, transitions, outcomes, and transition x outcome interactions on future decisions, we created vectors representing these variables. A trial was coded as +1 if the left port was chosen, and -1 if the right port was chosen. The correct action vector was assigned a value of 1 when an action was selected that led to a high probability of transitioning to a high reward probability state, and 0 otherwise. The outcome vector was assigned a value of 1 if the trial was rewarded, and 0 otherwise. The transition vector was assigned a value of 1 the rat experienced a common transition, and 0 otherwise. The transition x outcome interaction vector was obtained by multiplying the transition and outcome vectors. The following equation was used in the regression:

$$\log\left(\frac{p_{left}(t)}{1 - p_{right}(t)}\right) = \sum_{a=1}^{T}\beta_{choice}(a) * Choice(t-a) + \sum_{a=1}^{T}\beta_{correct}(a) * Correct(t-a) + \sum_{a=1}^{T}\beta_{transition}(a) * Transition(t-a) + \sum_{a=1}^{T}\beta_{outcome}(a) * Outcome(t-a) \sum_{a=1}^{T}\beta_{transition_outcome}(a) * Transition_outcome(t-a)$$

ACKNOWLEDGEMENTS

Funding: This work was supported by a JHU Kavli NDI Fellowship (YC), and National Institutes of Health grants F32DA054767 (EG), R01DA035943 (PHJ), and R01AA026306 (PHJ). **Author contributions:** EG, YC, and PHJ designed the experiments. EG, YC, LC, AD, CS, YW, ZC, NM, JM, and AN collected the data. EG and YC visualized and analyzed the data with input from PHJ. EG and YC prepared the manuscript with input from LC and PHJ.

REFERENCES

Akam, T., Costa, R., & Dayan, P. (2015). Simple Plans or Sophisticated Habits? State, Transition and Learning Interactions in the Two-Step Task. *PLoS Computational Biology*, 11(12), 1–25. https://doi.org/10.1371/journal.pcbi.1004648

Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R. F., … Costa, R. M. (2021). The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection. *Neuron*, 109, 149–163. https://doi.org/10.1016/j.neuron.2020.10.013

Anderson, E. M., Engelhardt, A., Demis, S., Porath, E., & Hearing, M. C. (2021). Remifentanil self-administration in mice promotes sex-specific prefrontal cortex dysfunction underlying deficits in cognitive flexibility. *Neuropsychopharmacology*, 46(10), 1734–1745. https://doi.org/10.1038/s41386-021-01028-z

Bakhti-Suroosh, A., Towers, E. B., & Lynch, W. J. (2021). A buprenorphine-validated rat model of opioid use disorder optimized to study sex differences in vulnerability to relapse. *Psychopharmacology*, *238*(4), 1029–1046. https://doi.org/10.1007/s00213-020-05750-2

Belin, D., & Everitt, B. J. (2008). Cocaine Seeking Habits Depend upon Dopamine-Dependent Serial Connectivity Linking the Ventral with the Dorsal Striatum. *Neuron*, *57*(3), 432–441. https://doi.org/10.1016/j.neuron.2007.12.019

Boulos, L. J., Nasseef, M. T., McNicholas, M., Mechling, A., Harsan, L. A., Darcq, E., … Kieffer, B. L. (2019). TouchScreen-based phenotyping: altered stimulus/reward association and lower perseveration to gain a reward in mu opioid receptor knockout mice. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-40622-6

Coffey, K. R., Nickelson, W. B., Dawkins, A. J., & Neumaier, J. F. (2023). Rapid appearance of negative emotion during oral fentanyl self-administration in male and female rats. *Addiction Biology*, *28*(12), 1–13. https://doi.org/10.1111/adb.13344

Colwill, R. M., & Rescorla, R. a. (1985). Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *Journal of Experimental Psychology: Animal Behavior Processes, 11*(4), 520–536. https://doi.org/10.1037/0097-7403.11.4.520

Corbit, L. H., Nie, H., & Janak, P. H. (2012). Habitual alcohol seeking: Time course and the contribution of subregions of the dorsal striatum. *Biological Psychiatry*, 72(5), 389–395. https://doi.org/10.1016/j.biopsych.2012.02.024

Corbit, L. H., Chieng, B. C., & Balleine, B. W. (2014). Effects of Repeated Cocaine Exposure on Habit Learning and Reversal by N-Acetylcysteine. *Neuropsychopharmacology*, *39*(8), 1893–1901. https://doi.org/10.1038/npp.2014.37

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Dickinson, A. (1985). Actions and Habits: The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *308*(1135), 67–78.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, 80(2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Garr, E., Padovan-Hernandez, Y., Janak, P. H., & Delamater, A. R. (2021). Maintained goal-directed control with overtraining on ratio schedules. *Learning and Memory*, *28*(12), 435–439. https://doi.org/10.1101/lm.053472.121

Giuliano, C., Puaud, M., Cardinal, R. N., Belin, D., & Everitt, B. J. (2021). Individual differences in the engagement of habitual control over alcohol seeking predict the development of compulsive alcohol seeking and drinking. *Addiction Biology*, *26*(6), 1–11. https://doi.org/10.1111/adb.13041

Groman, S. M., Massi, B., Mathias, S. R., Curry, D. W., Lee, D., & Taylor, J. R. (2019). Neurochemical and behavioral dissections of decision-making in a rodent multistage task. *Journal of Neuroscience*, *39*(2), 295–306. https://doi.org/10.1523/JNEUROSCI.2219-18.2018

Halbout, B., Hutson, C., Agrawal, S., Springs, Z. A., & Ostlund, S. B. (2024). Differential effects of acute and prolonged morphine withdrawal on motivational and goal-directed control over reward-seeking behaviour. *Addiction Biology*, *29*(5), 1–16. https://doi.org/10.1111/adb.13393

Hammerslag, L. R., Denehy, E. D., Carper, B., Nolen, T. L., Prendergast, M. A., & Bardo, M. T. (2021). Effects of the glucocorticoid receptor antagonist PT150 on stress-induced fentanyl seeking in male and female rats. *Psychopharmacology*, *238*(9), 2439–2447. https://doi.org/10.1007/s00213-021-05865-0

Hasz, B. M., & Redish, A. D. (2018). Deliberation and Procedural Automation on a Two-Step Task for Rats. *Frontiers in Integrative Neuroscience*, *12*(August), 1–19. https://doi.org/10.3389/fnint.2018.00040

Hölter, S. M., Tzschentke, T. M., & Schmidt, W. J. (1996). Effects of amphetamine, morphine and dizocilpine (MK-801) on spontaneous alternation in the 8-arm radial maze. *Behavioural Brain Research*, *81*(1–2), 53–59. https://doi.org/10.1016/S0166-4328(96)00043-5

Hynes, T., Fouyssac, M., Puaud, M., Joshi, D., Chernoff, C., Stiebahl, S., … Belin, D. (2024). Pan-striatal reduction in the expression of the astrocytic dopamine transporter precedes the development of dorsolateral striatum dopamine-dependent incentive heroin seeking habits. *European Journal of Neuroscience*, *59*(10), 2502–2521. https://doi.org/10.1111/ejn.16354

Jones, B. O., Cruz, A. M., Kim, T. H., Spencer, H. F., & Smith, R. J. (2022). Discriminating goal-directed and habitual cocaine seeking in rats using a novel outcome devaluation procedure. Learning and Memory, 29(12), 447–457. https://doi.org/10.1101/lm.053621.122

Korin, H. (1974). Comparison of psychometric measures in psychiatric patients using heroin and other drugs. Journal of Abnormal Psychology, 83(2), 208–212. https://doi.org/10.1037/h0036480

Kim, D., Park, G. Y., O′Doherty, J. P. , & Lee, S. W. (2019). Task complexity interacts with state-space uncertainty in the arbitration between model-based and model-free learning. *Nature Communications*, 10, 5738.

Kleber, H. D. (2007). Pharmacologic treatments for opioid dependence: Detoxification and maintenance options. Dialogues in Clinical Neuroscience, 9(4), 455–470. https://doi.org/10.31887/dcns.2007.9.4/hkleber

LaRocco, K., Villiamma, P., Hill, J., Russell, M. A., DiLeone, R. J., & Groman, S. M. (2024). Sex differences in oxycodone-taking behaviors are linked to disruptions in reward-guided, decision-making functions. *BioRxiv*. https://doi.org/10.1101/2024.04.09.587443

LeBlanc, K. H., Maidment, N. T., & Ostlund, S. B. (2013). Repeated Cocaine Exposure Facilitates the Expression of Incentive Motivation and Induces Habitual Control in Rats. *PLoS ONE*, 8(4). https://doi.org/10.1371/journal.pone.0061355

Little, K. M., & Kosten, T. A. (2023). Focus on fentanyl in females: Sex and gender differences in the physiological and behavioral effects of fentanyl. *Frontiers in Neuroendocrinology*, *71*(March), 101096. https://doi.org/10.1016/j.yfrne.2023.101096

Loh, E. A., Smith, A. M., & Roberts, D. C. S. (1993). Evaluation of response perseveration of rats in the radial arm maze following reinforcing and nonreinforcing drugs. *Pharmacology, Biochemistry and Behavior*, *44*(3), 735–740. https://doi.org/10.1016/0091-3057(93)90192-

Lyvers, M., & Yakimoff, M. (2003). Neuropsychological correlates of opioid dependence and withdrawal. *Addictive Behaviors*, *28*(3), 605–611. https://doi.org/10.1016/S0306-4603(01)00253-2

Ma, B., Mei, D., Wang, F., Liu, Y., & Zhou, W. (2019). Cognitive enhancers as a treatment for heroin relapse and addiction. *Pharmacological Research*, *141*(December 2018), 378–383. https://doi.org/10.1016/j.phrs.2019.01.025

Malone, S. G., Keller, P. S., Hammerslag, L. R., & Bardo, M. T. (2021). Escalation and reinstatement of fentanyl self-administration in male and female rats. *Psychopharmacology*, *238*(8), 2261–2273. https://doi.org/10.1007/s00213-021-05850-7

Miller, K. J., Brody, C. D., & Botvinick, M. M. (2016). Identifying Model-Based and Model-Free Patterns in Behavior on Multi-Step Tasks. *BioRxiv*, 096339. https://doi.org/10.1101/096339

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience*, *20*(9), 1269–1276. https://doi.org/10.1038/nn.4613

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values. *Psychological Review*, *126*(2), 292–311. https://doi.org/10.1037/rev0000120

Murray, J. E., Belin-Rauscent, A., Simon, M., Giuliano, C., Benoit-Marand, M., Everitt, B. J., & Belin, D. (2015). Basolateral and central amygdala differentially recruit and maintain dorsolateral striatumdependent cocaine-seeking habits. *Nature Communications*, 6, 1–9. https://doi.org/10.1038/ncomms10088

Nicolas, C., Zlbnik, N. E., Farokhnia, M., Lggio, L., Ikmoto, S., & Shaham, Y. (2022). Sex Differences in Opioid and Psychostimulant Craving and Relapse: A Critical Review. *Pharmacological Reviews*, *74*(1), 119–140. https://doi.org/10.1124/pharmrev.121.000367

Nordquist, R. E., Voorn, P., de Mooij-van Malsen, J. G., Joosten, R. N. J. M. A., Pennartz, C. M. A., & Vanderschuren, L. J. M. J. (2007). Augmented reinforcer value and accelerated habit formation after repeated amphetamine treatment. *European Neuropsychopharmacology*, 17(8), 532–540. https://doi.org/10.1016/j.euroneuro.2006.12.005

Sanchez-Roige, S., Ripley, T. L., & Stephens, D. N. (2015). Alleviating waiting impulsivity and perseverative responding by μ-opioid receptor antagonism in two inbred mouse strains. *Psychopharmacology*, *232*(8), 1483–1492. https://doi.org/10.1007/s00213-014-3786-9

Sebold, M., Deserno, L., Nebe, S., Schad, D. J., Garbusow, M., Hägele, C., … Huys, Q. J. M. (2014). Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology*, *70*(2), 122–131. https://doi.org/10.1159/000362840

Seip-Cammack, K. M., & Shapiro, M. L. (2014). Behavioral flexibility and response selection are impaired after limited exposure to oxycodone. *Learning and Memory*, *21*(12), 686–695. https://doi.org/10.1101/lm.036251.114

Tamargo, J. A., Campa, A., Martinez, S. S., Li, T., Sherman, K. E., Zarini, G., … Baum, M. K. (2021). Cognitive Impairment among People Who Use Heroin and Fentanyl: Findings from the Miami Adult Studies on HIV (MASH) Cohort. Journal of Psychoactive Drugs, 53(3), 215–223. https://doi.org/10.1080/02791072.2020.1850946

Thrailkill, E. A., Trask, S., Vidal, P., Alcalá, J. A., & Bouton, M. E. (2018). Stimulus control of actions and habits: A role for reinforcer predictability and attention in the development of habitual behavior. *Journal of Experimental Psychology: Animal Learning and Cognition*, *44*(4), 370–384. https://doi.org/10.1037/xan0000188

Towers, E. B., Setaro, B., & Lynch, W. J. (2022). Sex- and Dose-Dependent Differences in the Development of an Addiction-Like Phenotype Following Extended-Access Fentanyl Self-Administration. *Frontiers in Pharmacology*, *13*(March), 1–15. https://doi.org/10.3389/fphar.2022.841873

Towers, E. B., Setaro, B., & Lynch, W. J. (2023). Estradiol Enhances the Development of Addiction-Like Features in a Female Rat Model of Opioid Use Disorder. *Neuroendocrinology*, 113(11), 1099–1111. https://doi.org/10.1159/000529997

Vandaele, Y., & Ahmed, S. H. (2021). Habit, choice, and addiction. *Neuropsychopharmacology*, *46*(4), 689–698. https://doi.org/10.1038/s41386-020-00899-y

Vandaele, Y., & Janak, P. H. (2018). Defining the place of habit in substance use disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry,* 87, 22–32. https://doi.org/10.1016/j.pnpbp.2017.06.029

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., … Bullmore, E. T. (2015). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry*, *20*(3), 345–352. https://doi.org/10.1038/mp.2014.44

Wheeler, A. R., Truckenbrod, L. M., Cooper, E. M., Betzhold, S. M., Setlow, B., & Orsini, C. A. (2023). Effects of fentanyl self-administration on risk-taking behavior in male rats. Psychopharmacology, 240(12), 2529–2544. https://doi.org/10.1007/s00213-023-06447-y

Wakhlu, A., Engelhardt, A., Anderson, E. M., Grafelman, E., Ouimet, A., & Hearing, M. C. (2023). Adolescent morphine exposure impairs cognitive flexibility in female but not male mice. BioRxiv. https://doi.org/10.1101/2023.04.09.536177

Wang, W., Xie, X., Zhuang, X., Huang, Y., Tan, T., Gangal, H., … Wang, J. (2023). Striatal μ-opioid receptor activation triggers direct-pathway GABAergic plasticity and induces negative affect. *Cell Reports*, *42*(2), 112089. https://doi.org/10.1016/j.celrep.2023.112089

Zapata, A., Minney, V. L., & Shippenberg, T. S. (2010). Shift from Goal-Directed to Habitual Cocaine Seeking after Prolonged Experience in Rats. *The Journal of Neuroscience*, *30*(46), 15457–15463. https://doi.org/10.1523/JNEUROSCI.4072-10.2010

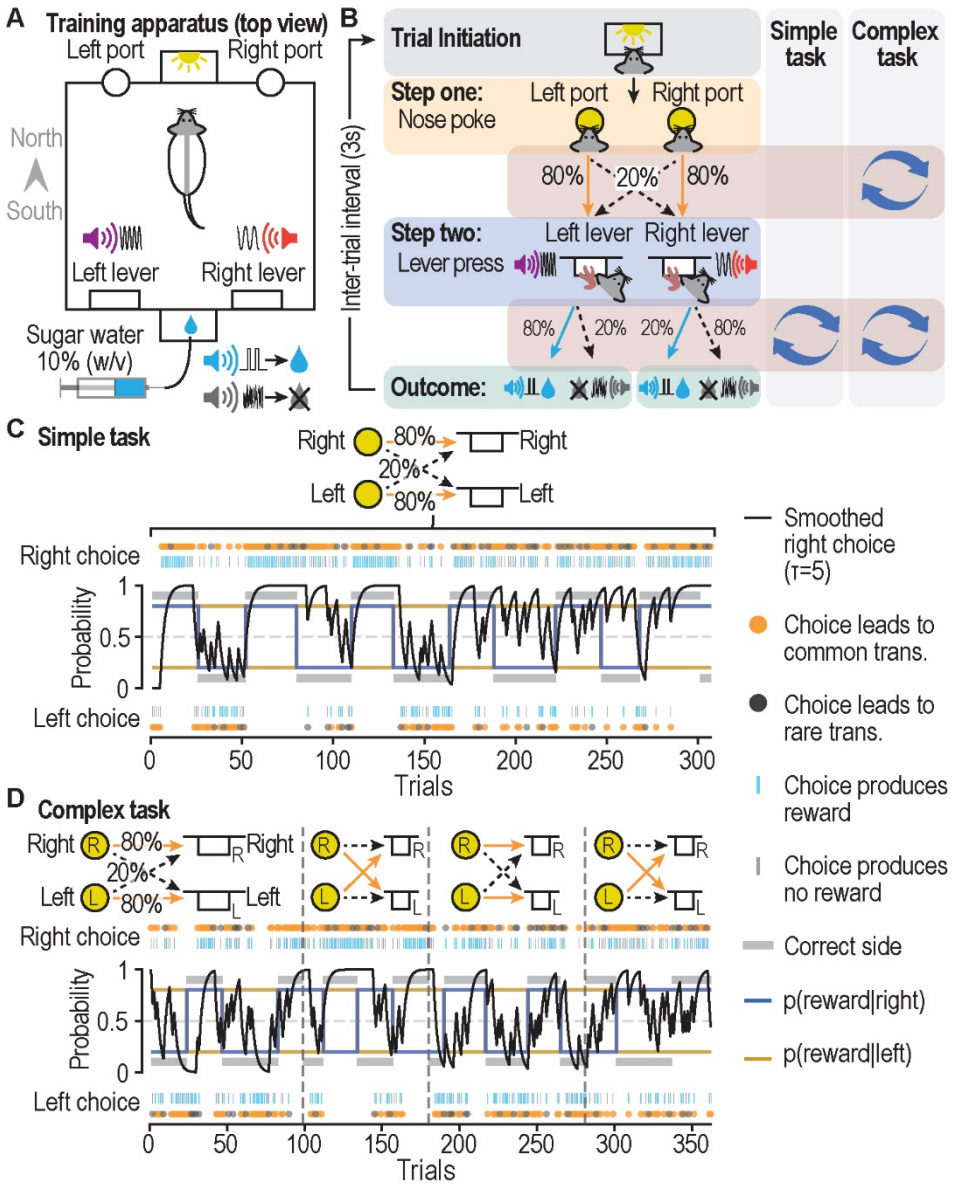

Figure 1. (A) Schematic of behavioral apparatus, consisting of one magazine for center fixation and two nose-ports on the north side, and one magazine for reward and two retractable levers on the south side. Two speakers are mounted on the south wall. (**B**) Structure of a single trial in the two-step task. After the rat nose-pokes, the magazine on the north wall illuminates, the choice ports light up, and the rat is free to make a decision by nose-poking into one of the ports. The chosen port probabilistically leads to the insertion of a lever on the south wall, accompanied by an auditory tone indicating the lever's location. Once the rat presses the extended lever, a reward (a drop of 10% sucrose) is delivered to the magazine on the south wall. Rewarded deliveries are signaled by a clicker, while reward omissions are accompanied by white noise. The simple two-step task contains only reward probability reversals, whereas the complex task contains reversals for reward probabilities and transition probabilities. (**C,D**) Example behavior session of simple task (C) and complex task (D).

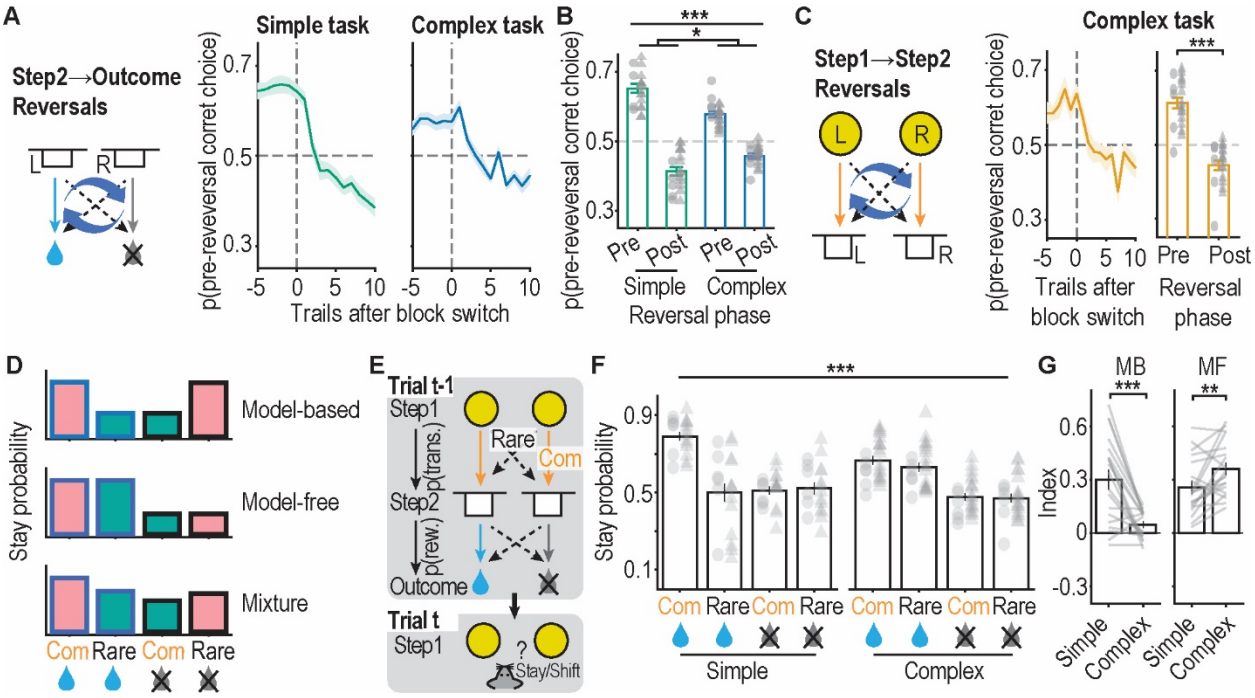

Figure 2. (A) Probability of choosing the nose-poke that leads to the higher pre-reversal probability of reward (correct choice) as a function of time relative to the reward probability reversal for each task type (simple vs. complex). Data represent means across rats, with shaded areas indicating SEMs. (**B**) Correct choice probabilities from A are averaged over 5 trials before reversal ('pre') and compared to probabilities averaged over trials 6-10 after reversal ('post'). (**C**) Probability of choosing the correct choice relative to the transition probability reversal for the complex task (left panel). The right panel shows averaged pre-reversal correct choice probabilities (5 trials before reversal) compared to post-reversal (trials 6-10 after reversal). (**D**) Likelihood of repeating actions (stay probabilities) is affected by previous trial outcome (water drop icon for reward, and water drop icon plus a cross for omission) and transition [common (com), rare]. (**E**) Model-based, model-free, and mixture of both show distinct pattern of conditional stay probabilities. (**F**) Conditional stay probabilities in simple and complex tasks. (**G**) MF and MB indices derived from the stay probabilities in F. Bars represent means across rats, error bars show SEMs for B,C,F, and G. Individual male rats shown as dots and individual female rats as triangles for B, C, and F. Lines in G indicate individual rats including both male and female.

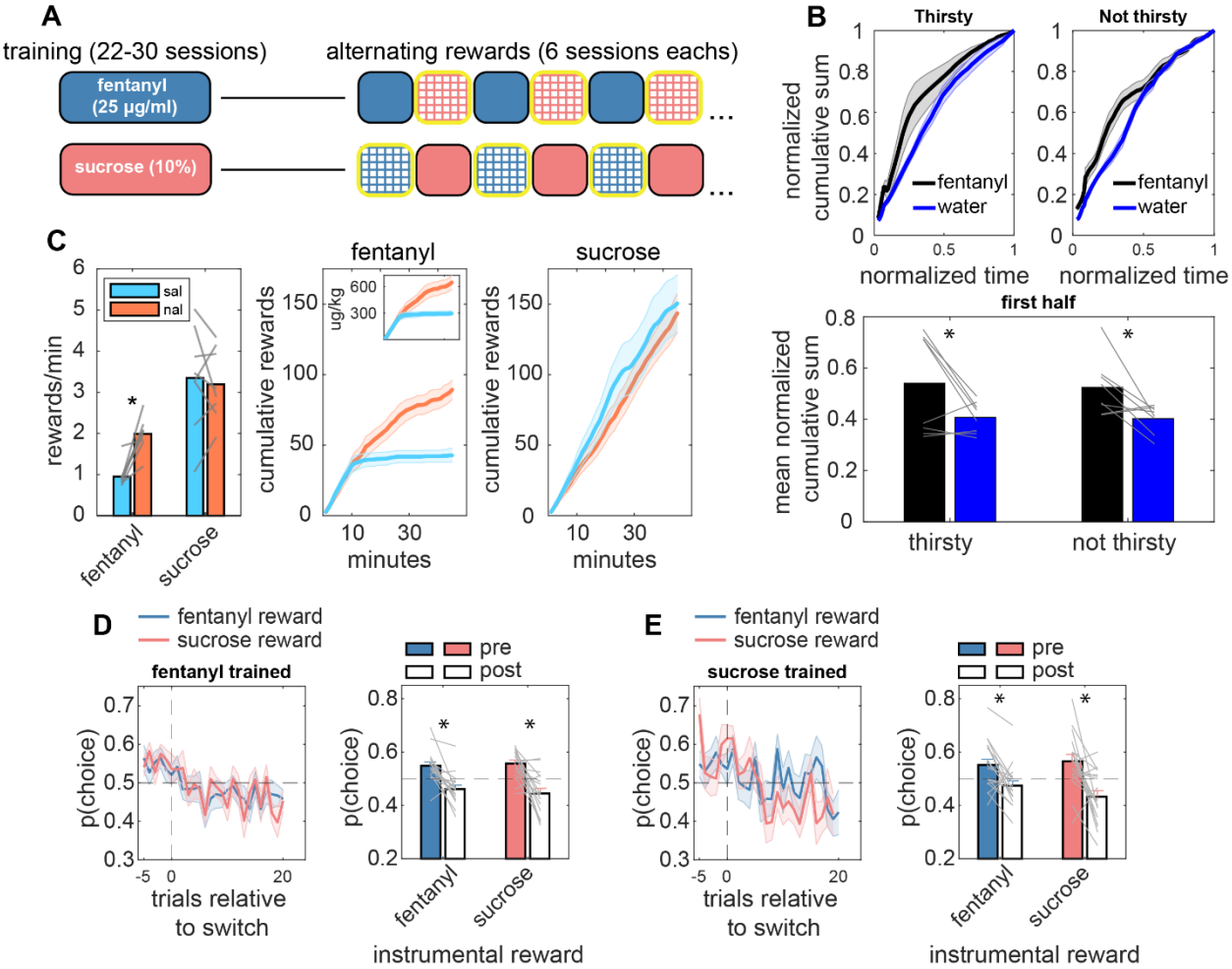

Figure 3. (A) Illustration of experimental design. Two groups of rats were trained to earn either oral fentanyl or sucrose on the two-step task for many sessions, and then experienced alternating sessions of fentanyl and sucrose. Alternating sessions with the reward type not experienced during training took place in contexts with a different scent and floor texture. **(B)** Normalized cumulative sum of fentanyl and water rewards as a function of normalized session time as rats pressed a lever for liquid rewards under thirsty and non-thirsty conditions. Cumulative sums were normalized in the range of 0-1 and then interpolated. **(C)** Left: Mean reward rates while rats pressed a lever on an FR5 schedule for either liquid fentanyl or sucrose after receiving IP injections of saline or naltrexone. Grey lines are individual rats. Middle: Cumulative fentanyl rewards earned across sessions. Inset shows the cumulative intake. Right: Cumulative sucrose rewards earned across sessions. **(D)** Left: Probability of choosing the nose-poke that leads to a higher pre-reversal probability of reward as a function of the relative time from reward probability reversal for each training group (fentanyl vs. sucrose) and each instrumental reward (fentanyl vs. sucrose) for rats given extensive training with fentanyl. Data show means across rats, and shaded bounds are SEMs. Right: Nose-poke choice probabilities were averaged over trials 1-5 before reversal ('pre') and compared to choice probabilities averaged over trials 15-20 after reversal ('post'). Bars show means across rats, error bars show SEMs, and grey lines are individual rats. **(E)** Same as panel D but for the rats given extensive training with sucrose.

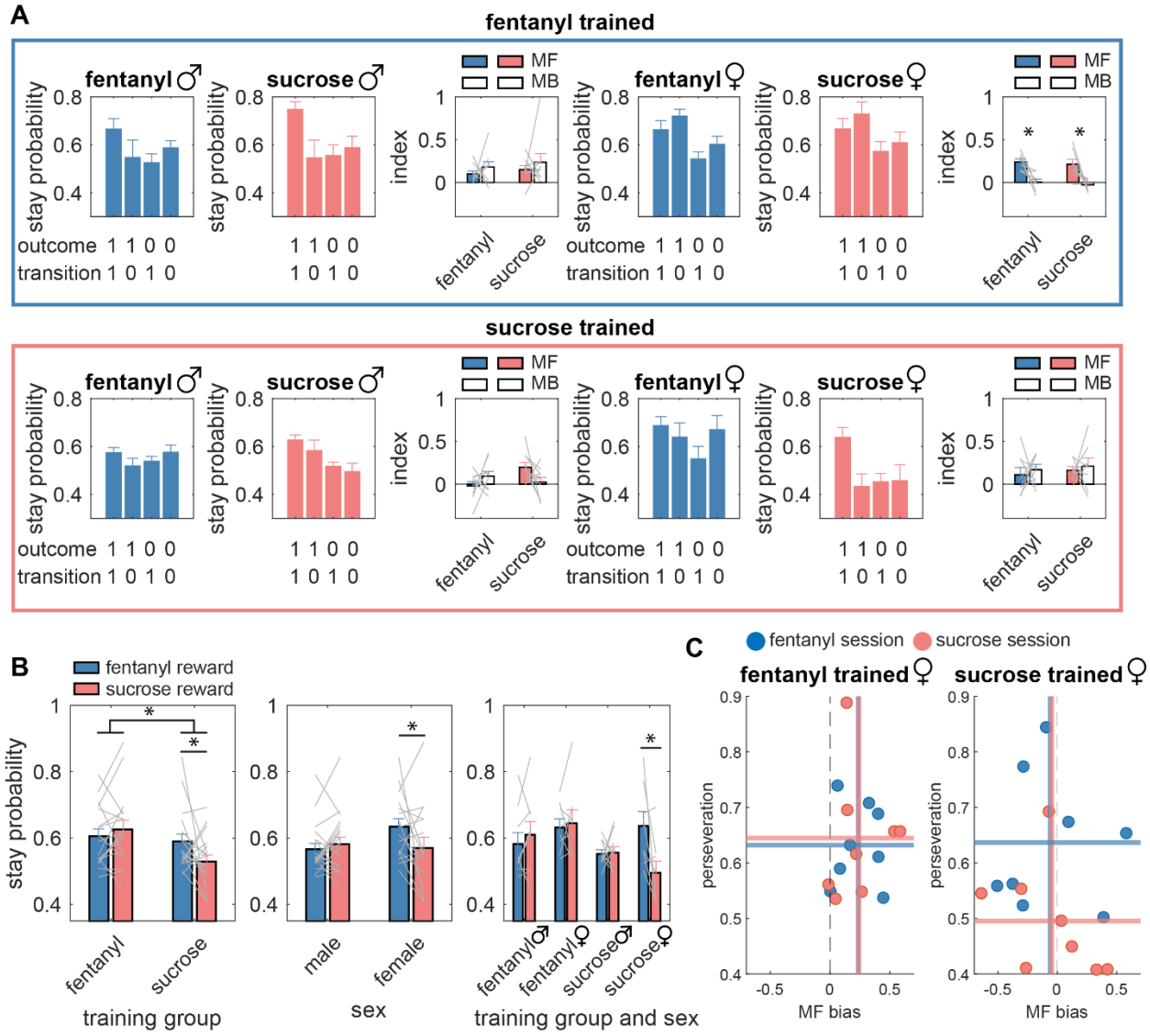

Figure 4. (A) Stay probabilities plotted as a function of the previous trial outcome (1 = reward, 0 = omission), transition (1 = common, 0 = rare), training group (fentanyl vs. sucrose), instrumental reward (fentanyl vs. sucrose), and sex (male vs. female). MF and MB indices derived from the stay probabilities are also shown. Bars show means across rats, error bars show SEMs, and grey lines are individual rats. **(B)** Stay probabilities averaged over trial-level variables (i.e. outcome and transition) and plotted as a function of training group and instrumental reward (left), sex and instrumental reward (middle), and all three variables combined (right). Bars show means across rats, error bars show SEMs, and grey lines are individual rats. **(C)** Summary of main findings. Perseveration, defined as the mean stay probability, is plotted against MF bias, defined as MF – MB index. Fentanyl trained females (left) showed a high level of perseveration and a high MF bias during fentanyl and sucrose sessions. Sucrose trained females (right) showed a high level of perseveration specifically during fentanyl sessions, but not MF bias. Each data point is an individual rat.

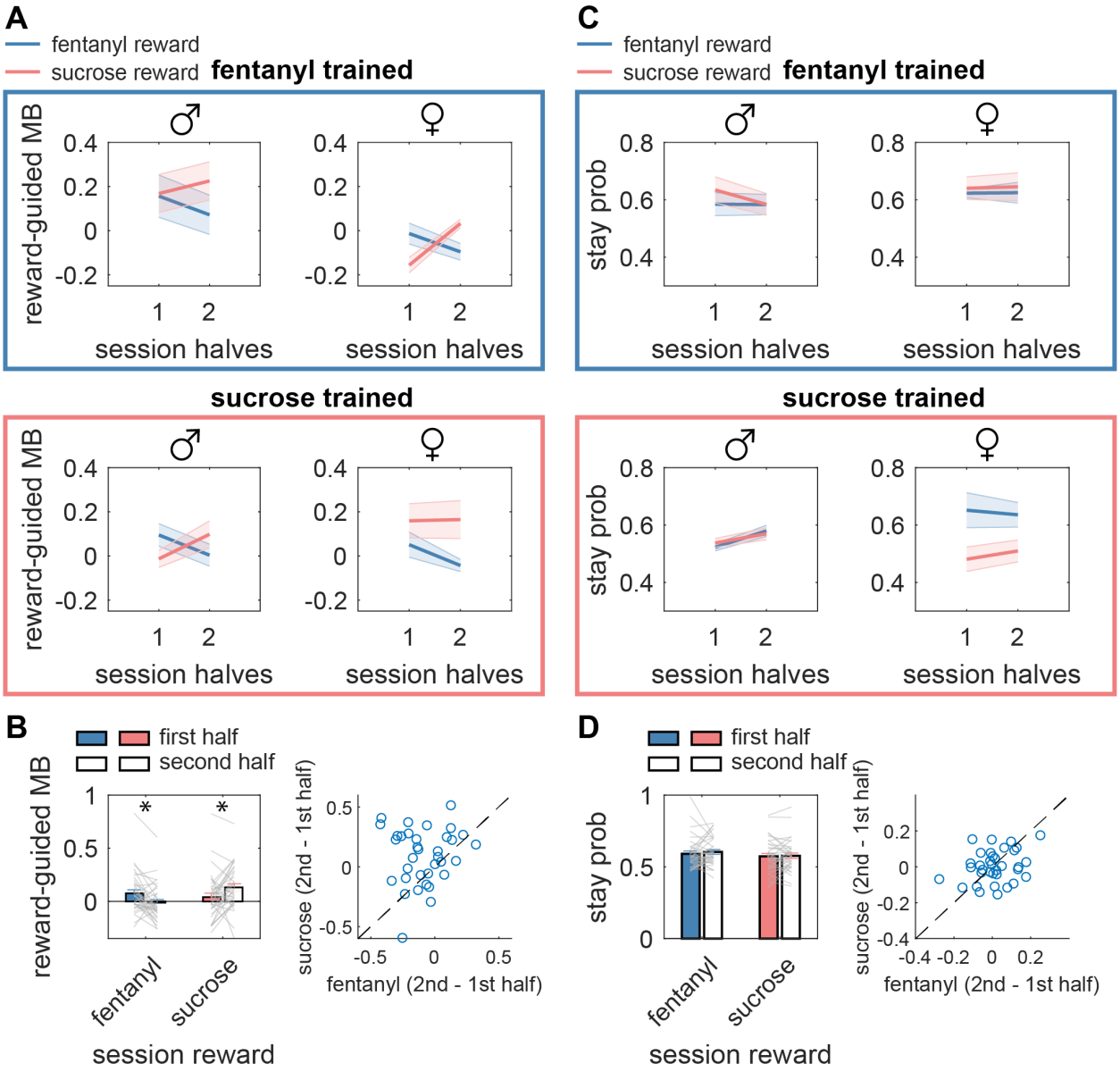

Figure 5. (A). Reward-guided MB indices, defined as the difference in the stay probabilities between reward-common and reward-rare trials, are shown as a function of session half separately for males and females trained with fentanyl (top) or sucrose (bottom). Lines are means across rats, and shaded bounds are SEMs. **(B)** Summary of main finding. Reward-guided MB indices are plotted as a function of instrumental reward and session half for all rats (left). The difference between reward-guided MB choice during the first and second halves of the sessions is plotted for individual rats during sucrose and fentanyl sessions. **(C)** Stay probabilities averaged across trial-level variables (i.e. outcome and transition) are plotted over sessions halves separately for instrumental reward, training group, and sex. **(D)** Summary of perseveration scores over time. Stay probabilities are plotted as a function of instrumental reward and session half for all rats (left). The difference between perseveration scores during the first and second halves of the sessions is plotted for individual rats during sucrose and fentanyl sessions.

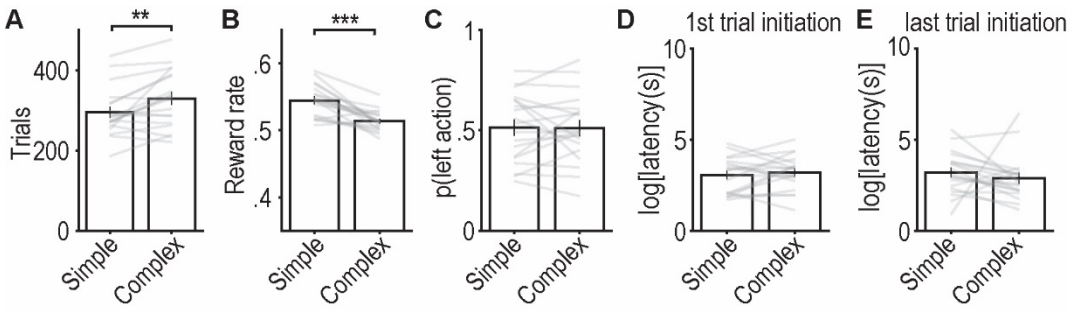

Figure S1. (A) Mean number of trials per session performed by rats in the simple task compared to the complex task. **$p < 0.01$. **(B)** Mean reward rate, defined as the ratio of rewarded trials to total trials, in the simple task compared to the complex task. ***$p < 0.001$. **(C)** Mean probability of choosing the left port in the simple task compared to the complex task. **(D)** Mean log-transformed latency to initiate the first trial of a session in the simple task compared to the complex task. **(E)** Average log-transformed latency to initiate the last trial of a session in the simple task compared to the complex task**.** Bars show means across rats, error bars show SEMs, and grey lines are individual rats. Paired two-tailed t-tests for A and B.

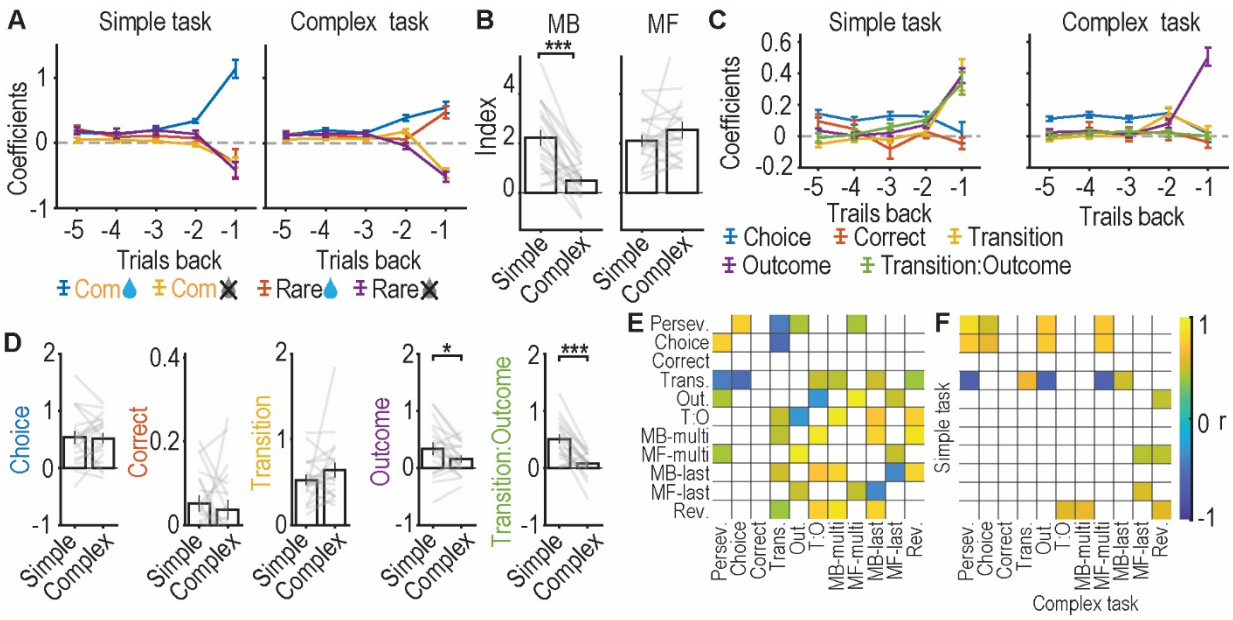

Figure S2. (A) Regression coefficients as a function of the past 5 trials, relative to choice, previous trial outcome, transition, and task type. Lines represent means across rats, and error bars indicate SEMs. (**B**) Model-free (MF) and model-based (MB) indices derived from the 5-trial back coefficients shown in A. Bars represent means across rats, error bars indicate SEMs, and grey lines show individual rats. *** $p < 0.001$. (**C**) Regression coefficients as a function of the past 5 trials, relative to choice, previous trial choice direction, correctness, outcome, transition, outcome-transition interaction, and task type. Lines represent means across rats, and error bars indicate SEMs. (**D**) Sum of each coefficient from (C) across the past 5 trials for the simple task compared to the complex task. Bars represent means across rats, error bars indicate SEMs, and grey lines show individual rats. * $p < 0.05$, *** $p < 0.001$. (**E**) Correlation among behavioral measures from Figure S1 and within this figure, irrespective of task type. Choice perseveration (Persev.) was calculated as the average stay probability across each outcome-transition combination. (**F**) Correlation among behavioral measures in E between the simple and complex tasks. Significant ($p < 0.05$) correlations are shown in non-white colors, with the correlation strength and direction indicated by the color bar for both E and F. Paired two-tailed t-tests were used for comparisons in B and D.

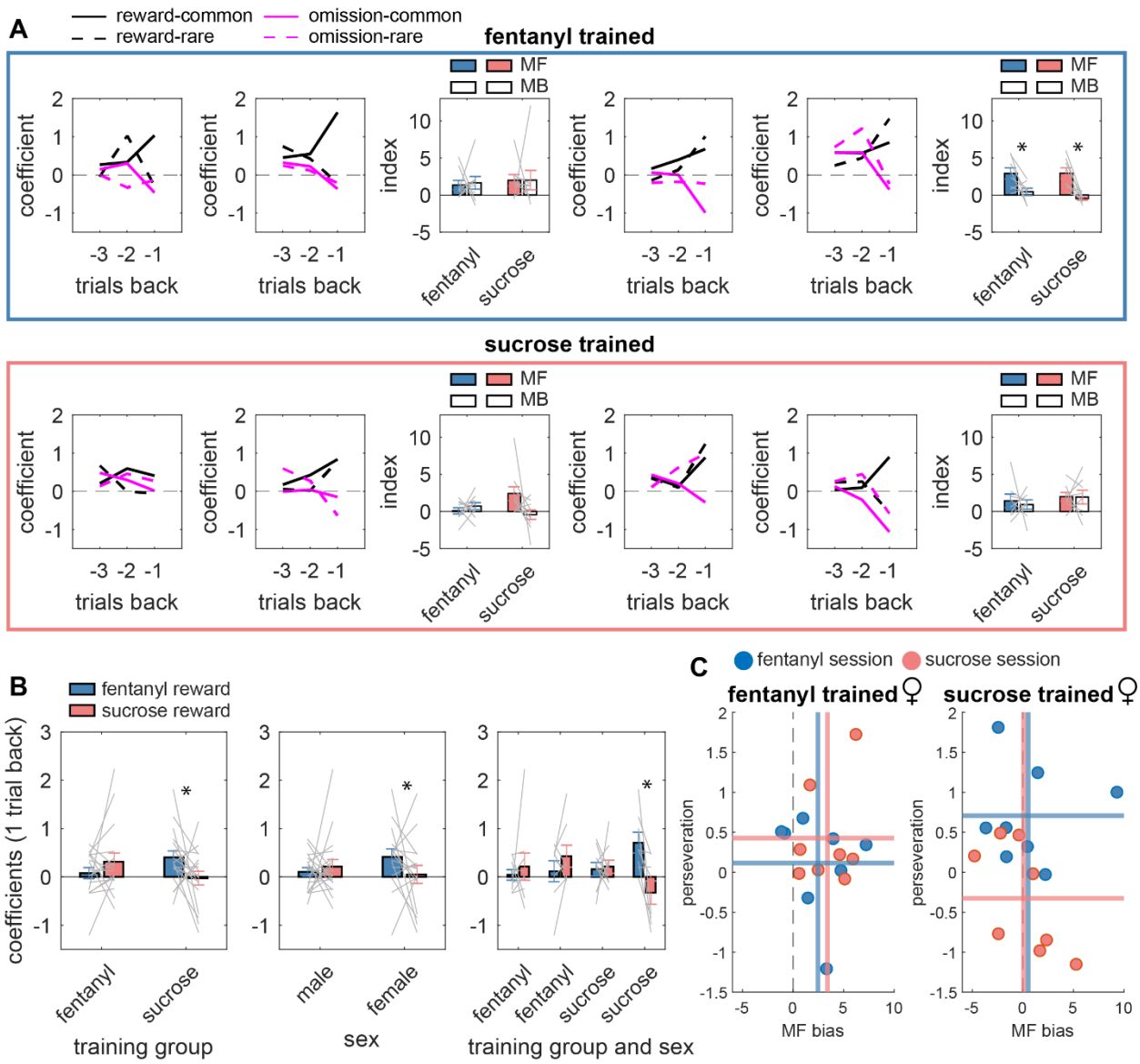

Figure S3. (A) Regression coefficients plotted as a function of past trials relative to choice, the previous trial outcome, transition, training group, instrumental reward, and sex. MF and MB indices derived from the coefficients one trial back are also shown. Lines and bars show means across rats, error bars show SEMs, and grey lines are individual rats. **(B)** Regression coefficients one trial back averaged over trial-level variables (i.e. outcome and transition) and plotted as a function of training group and instrumental reward (left), sex and instrumental reward (middle), and all three variables combined (right). Bars show means across rats, error bars show SEMs, and grey lines are individual rats. **(C)** Perseveration, defined as the mean regression coefficient one trial back, is plotted against MF bias, defined as MF – MB index. Each data point is an individual rat.

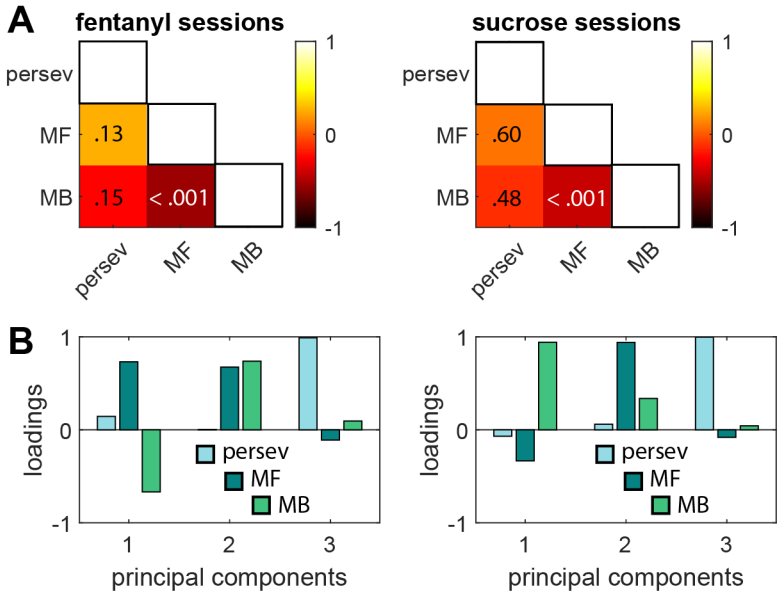

Figure S4. (A) Heat map showing correlations between perseveration scores, MF indices, and MB indices separately for fentanyl and sucrose sessions. Colors represent correlation coefficients, and text inside matrix cells are *p* values. **(B)** Principle component analysis on perseveration scores, MF indices, and MB indices separately for fentanyl and sucrose sessions. Perseveration loads overwhelmingly on PC3 with negligent contributions from MF and MB indices.

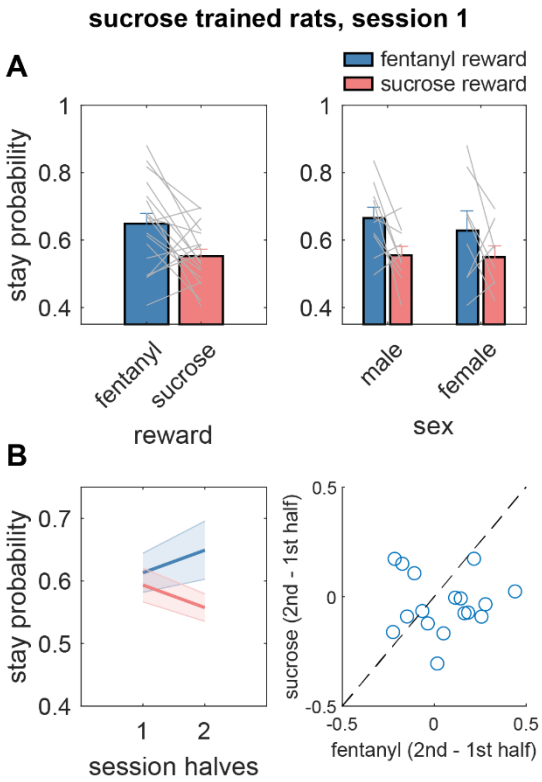

Figure S5. (A) Mean stay probabilities from the first fentanyl session, and the subsequent sucrose session, for rats previously given extensive training with sucrose. Data are shown collapsed across (left) and divided by (right) sex. **(B)** Mean stay probabilities divided by session halves and reward type (left). The difference between halves is shown for individual rats for fentanyl and sucrose sessions (right).