Peer Review File

Manuscript Title: TRANSFORMATIONS AND FUNCTIONS OF NEURAL REPRESENTATIONS IN A SOCIAL BEHAVIOR NETWORK

Editorial Notes:

Redactions – unpublished data

Parts of this Peer Review File have been redacted as indicated to maintain the confidentiality of unpublished data.

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

In "BNST promotes regional male bias within a female-biased circuit controlling social behavior in male mice" Yang and Anderson start with knowledge from a previous Shah lab study of bulk activity of the AB+ BNSTpr neurons which concluded that they encode representations of gender in the male brain in an intensity-dependent manner. The present study seeks to dissect the role of BNST-ESR neurons (a subset of AB+ neurons) that project to either the MPOA or the VMHvl using single cell imaging. The approach is well used and the depth of the analysis performed is complete and convincing, leading to technically rigorous experiments. However, the difference between what was previously shown and what is now understood is subtle and less convincing. Moreover, neither the previous or the current study convince me that these neurons are actually encoding sex representations, largely because in the previous study they were not found to encode gender in the female brain (and it was not studied here). Is it likely that the female brain has evolved an entirely different mechanism to encode such a key percept? The analysis in this study nicely supports that they are responding to an analog comparison but it could be another function, such as motivation, that drives their activity. Though it may be intuitive to think that they are encoding sex and is supported by the PC and in silico analysis, it is not empirically shown by experimental means. It is correlative. The author's statement that it is "paradoxical that silencing of BNSTpreEsr1 neurons had no effect on our ability to decode intruder sex from population activity in either the MPOA or VMHvl" indicates that the data, model, and hypothesis are not yet fully aligned. The other major finding from this study is the flexibility of the neural code downstream of the BNST. This part of the conclusion echos back to earlier work from the Anderson lab finding scalable control of mounting and attack in the VMH, and work from the Lin lab has shown heterogeneity in the VMH-Esr population that correlates with promoting different social behaviors (Hashikawa, 2017). This extends some of that previous work by focusing on the function of the BNST to gate differential activity, but there is no mechanistic understanding of how these neurons are dynamically recruited or how their addition alters natural social behavior. Though I understand the meaning of the title, I do not agree that the statistical difference of sampling of several 100 neurons in one area of the brain should be interpreted as functional male or female bias. Though the experiments in this study are elegant, they do not serve to clarify the significance of the BNST in the social behavior circuit and the representation of sex in the brain.

Additional concerns:

1) From fig 3 onwards, where the chemoinhibition of BNST Esr1 neurons is being used to draw conclusions about the information content in the VMHvl Esr1 and the MPOA Esr1 neurons, the stated hypothesis is "that inhibiting BNSTprEsr1 neurons should alter sex representations in MPOAand VMHvl." To establish this, it would be clearer to inhibit BNST Esr1 neurons, and then image neurons in the VMHvl or MPOA that express Esr1 AND are downstream of the BNST Esr1 (or BNST) neurons. In the absence of this kind of experiment, it is difficult to formally state that the BNST- VMHvl or BNST-MPOA projection has anything to do with the distortions in representations at the VMHvl or MPOA.

2) The above concern is also applied to behavior. When the BNST-Esr to -VMH-Esr or to MPOA-Esr is specifically silenced, what is the effect on behavior?

3) The rationale to focus on the Esr population is not clear. Are the ESR negative subset of AB neurons not responding to male and female cues or projecting to the MPOA and VMH? If not, what is their role in the model?

4) Many of the BNST-Esr+ neurons are not active in the presence of either a male or female (Figure 2). Are these neurons also projecting to the MPOA or VMH? The use of mating and aggression as a behavioral proxy for sex identification is not very granular. Maybe the BNST contributes to other aspects of social motivation that result in the phenotypes in figure 1. Do the authors have another measure for sex recognition that could support the functional conclusions of the Esr neurons driving sex recognition?

5) Are all of the BNST-Esr neurons that project to the two studied targets inhibitory? If not, this could confound the model.

6) Is it possible to manipulate the sensory signals so that BNST and VMH activity can be monitored as they mis-identify the sex of the partner? A gender illusion? This would control for all other aspects of behavior and neural activity.

7) The model (EDF8) suggests that absolute incoming sensory activity is weighed, more female activity = mating, more male activity = aggression. Can you manipulate this by allowing your subjects to interact with multiple females and a single male simultaneously? It looks like the BNST activity persists as long as a female is present (fig 2j). Is this correct? If the subject first interacts with females and then a male is also added subsequently, does the simultaneous presence of a male alter the balance of activity in the VMH? (In both cases, I expect the presence of a male would promote aggression, but the representation of male and female should not change.)

8) Figure 4a, it appears that with CNO (BNST-) the female responding cells are largely spatially segregated from the male responding cells and about half of the male responding cells remain male responding. Are these two populations, those that switch vs retain sex tuning without BNST input different molecular subsets of the VMHvl-ESR population? On repeated trials, is the same neuron able to switch sometimes and remain stable other times, or are they set to be either flexible or fixed?

Minor Concerns:
1) I cannot find CNO only controls on neural activity and behavior, and some quantification/analysis for the silencing of BNST-Esr expressing hm4di neurons.
2) The term chemoscope doesn't seem to add much, it is fundamentally chemogenetics and regular miniscope imaging combined.
3) Fig 2: Representative images of GCaMP infections in BNSTEsr1 neurons will be useful.
4) Fig 3: Representative images of DREADDs infections in BNSTEsr1 neurons + GCaMP in MPOA/VMHvl Esr1 neurons will be useful.
5) Ext Fig 4 : Why are error bars missing in k?

Referee #2 (Remarks to the Author):

Neural circuits that mediate social behaviors like mating and aggression are not well understood. This paper is a technical tour-de-force, using in vivo calcium imaging in several limbic system nuclei of awake mice to provide single neuron representations during male and female social encounters. This study significantly advances our understanding of sensory coding in the BNST (a limbic system structure that receives chemosensory inputs), revealing that different neurons are tuned to male and female sensory cues, with smaller groups of neurons activated during behavioral displays rather than by sensory cues per se. Prior studies involving fiber photometry in this region were important, claiming a female-bias in the overall response, but as they lacked single neuron resolution, they missed the presence of male-selective neurons, which are sparser, as well as the distribution of neurons tuned to sensory vs motor actions. Moreover, Yang and Anderson used chemogenetics to show that BNST inputs re-shape sex representations in a downstream hypothalamic nucleus (the VMH). I am enthusiastic about this study, but also note that some additional controls are needed to validate claims and that some questions remain related to the mechanism underlying the interesting BNST-VMH transformation.

1. It is interesting that the VMH and MPOA still display sex-biased responses after chemogenetic inhibition of the BNST. This could be due to factors discussed, such as a role for another brain region, but also could be due to incomplete BNST inhibition in chemogenetic experiments, related to either (1) the extent of coverage of the large BNST area by AAV injections, and 2) the Cre lines used. For (1), what % of Cre-expressing BNST cells are labeled and silenced by AAV injections, and for (2) are there sex-selective BNST neurons that do not express aromatase and/or ESR and could be contributing to downstream representations?

2. The authors should ensure that AAV injection in the BNST does not label neurons in the VMH or MPOA (for example by retroactive labeling of neurons providing feedback control), which could confound interpretations.

3. The authors use a nicely comprehensive set of stimuli in Figure 2j, and showed that BNST response variance due to intruder sex was larger than the associated behavior. It would be worth discussing these observations in the context of the lab's previous work on VMH/MPOA. Is there an increased response variance due to behavior in the VMH and MPOA, suggesting further input transformation as information moves to the hypothalamus?

4. The authors focus here on neuronal representations in male mice; do similar sex biases in BNST and VMH exist in female mice?

5. Both the title and last sentence of the abstract should be edited for clarity to make the manuscript more accessible to a general audience. (I would go with something like 'Transformations of sex representations in the ascending limbic system' but of course this is just a suggestion!)

6. For ED 4b, it seems like activity is synchronized to particular events in the time series- if true, it would be helpful to provide annotation of whether such synchronized events correspond to sniffing or other social episodes.

Referee #3 (Remarks to the Author):

Yang & Anderson demonstrate that chemogenetic silencing of BNSTprEsr1 alters sexual behavior and aggression. Optogenetic silencing of this same neuronal population showed that activity of BNSTprEsr1 neurons is required for the transition from appetitive to consummatory social behaviors towards both sexes (i.e. initiation and duration of aggression and sexual behavior).
These findings (reported already in previous studies, as noted by the authors) set the foundation to determine how sex is presented in this neuronal population. The authors performed calcium imaging of the BNSTprEsr1 neuronal population in sexually experienced male mice during social interaction with an intruder male or female, using a miniature head-mounted microscope, and specifically monitored the decoding of the intruder's sex, at a single cell level. They identified subpopulations that were female preferring or male preferring, with or without physical interaction with the intruders, suggesting that sex is represented by population coding.
Next, they combined chemogenetic silencing of BNSTprEsr1 neurons, with microendoscopic imaging of VMHvlEsr1 or MPOAEsr1 neurons expressing. They revealed that such manipulation led to a decrease in the neuronal response to male intruders in both MPOA and VMHvl neurons, and an increase in the response to female intruders in VMHvl. Moreover, they revealed that silencing BNSTprEsr1 neurons inverted the 2:1 ratio of male- to female-preferring units in VMHvl, to the female-dominant ratio seen in MPOA.
The main novel findings and conclusion of the authors is that "the activity of BNSTprEsr1 neurons is not required for the coding of intruder sex identity by MPOAEsr1
and VMHvlEsr1 neurons. Rather, it is required to invert, in VMHvl, the female bias in population representations of intruder sex seen in BNSTpr, MPOA and MeApd, to a male bias".

The set of methodologies used is impressive, contain appropriate controls and the data is analyzed

well. No doubt it will provide a great dataset for better understanding of how sex-specific stimuli are encoded in neuronal networks, at single cell resolutions.
However, the neuroimaging-chemogenetic data set (the main part which provide novel findings), although interesting, is purely descriptive and missing any mechanistic explanation (beyond the proposed hypothesis). Namely, how does sex-biased neuronal coding in the MPOA/VMH/BNST (or the altered sex-bias in the VMH following chemogenetic silencing) encode sex-typical stimuli and control different reproductive behaviors (mating and aggression) towards males and females? As stated by the authors themselves in the discussion, the manuscript does not provide any experimental data to answer whether and how the changes in the ratio of female-male responsive neurons induced by silencing of BNSTprEsr1 are required for sex discrimination, sex-typical sexual/aggressive behavior, or any other phenotype. Moreover, it is essential to confirm the findings with additional complementary manipulations such as chemogenetic or optogenetic activation of the same neuronal population.

Additional concerns:
1. The use of restrained individuals as social stimui (first set of experiment) is very problematic, as it may trigger a massive stress response in both mice. If the authors wish to examine the response to a conspecific mouse without enabling attack or mount responses, they can present the conspecific separated by a perforated barrier, or use a stimuli such as urine or soiled bedding.

2. The link between the behavioral and neural effects of the chemogenetic silencing is poorly explained, thus it is not clear what is the biological/functional significance of the neural findings.

3. In continuous to the prior comment, did the authors notice any sex-reversed behaviors during the silencing period (i.e. attack of subject females or sexual behaviors towards males)? These behavior should be quantified for pre-CNO and CNO segments.

4. Also, is the silencing effect reversible? What would happen in optogenetic / chemogenetic activation of these neurons?

5. In order to support their sex discrimination claim, the authors need to conduct a separate discrimination assay, during BNSTprEsr1 silencing.

6. Do the response characteristics of specific neurons remain stable for days/weeks?

Author Rebuttals to Initial Comments:

Nature 2021-03-04882 Yang et al. Point-by-point response

Reviewer #1

We thank this reviewer for their incisive and helpful questions and comments. While it was not possible to perform every experiment requested, both for technical reasons and because of the contraction of our mouse colonies necessitated by the COVID pandemic, we have done our best to provide new data to address them.

Comment 1. "neither the previous or the current study convince me that these neurons are actually encoding sex representations, largely because in the previous study they were not found to encode gender in the female brain (and it was not studied here)."

> Response: We appreciate the reviewer's comment, but find it puzzling. We do not understand why evaluating the function of a neuronal population in males is dependent on knowing what a similar population does in females (or vice-versa), especially when the circuit we are studying is well-known for sexual dimorphisms in structure and function. Perhaps the reviewer thinks that by "sex representations" we mean a representation of the animal's OWN sex? If so, that is a misunderstanding: we meant the representation of the sex of a conspecific intruder, either male or female (see Remedios et al. (2017) *Nature* **550**:388-392). In any case, as detailed below our new data indicate that while BNSTpr may contain a neural representation of intruder sex (or a sex-specific internal state), it is not required functionally in males to identify and distinguish male from female conspecifics. This in turn argues that the requirement for BNSTpr in mounting and attack behavior is not an indirect consequence of deficient sex identification. These new findings have substantially changed our view of BNSTpr function, as explained in our responses below and in the revised manuscript.

Comment 2. "The analysis in this study nicely supports that they are responding to an analog comparison, but it could be another function, such as motivation, that drives their activity. Though it may be intuitive to think that they are encoding sex and is supported by the PC and in silico analysis, it is not empirically shown by experimental means. It is correlative."

Response: The reviewer is correct that we cannot distinguish a function in encoding sex from encoding a motivational state that is closely associated with the intruder's sex. Our experiments using functional silencing demonstrate that BNSTpr^{Esr1} neurons are necessary for the transition from sniffing to consummatory behavior (mating or fighting).

We have now revised our manuscript to indicate that BNSTpr could encode either intruder sex, or a motivational state that is strongly correlated with intruder sex, on **pg. 10 and 12.**

Comment 3. "From fig 3 onwards, where the chemo-inhibition of BNST Esr1 neurons is being used to draw conclusions about the information content in the VMHvl Esr1 and the MPOA Esr1 neurons, the stated hypothesis is "that *inhibiting BNSTprEsr1 neurons should alter sex representations in MPOA and VMHvl*." To establish this, it would be clearer to inhibit BNST Esr1 neurons, and then **image neurons in the VMHvl or MPOA that express Esr1 AND are downstream of the BNST Esr1 (or BNST) neurons**. In the absence of this kind of experiment, it is difficult to formally state that the BNST-VMHvl or BNST-MPOA projection has anything to do with the distortions in representations at the VMHvl or MPOA".

Response: The experiment the reviewer suggests (**bolded** above) is not technically feasible at present. To our knowledge, there is no method that will efficiently transfer GCaMP anterogradely from a pre-synaptic neuron to its post-synaptic target while maintaining the viability of the latter cells for imaging. However, we now present new data in **ED Fig. 2** showing that silencing of BNSTpr^{Esr1} terminals in VMHvl and MPOA blocks attack and mating, respectively (but not vice-versa). We also present new data showing that the female-biased representations of intruder sex are present in *BNSTpr neurons that project to MPOA and VMHvl* (**Extended Data Fig. 9**). Together these data make it highly likely that the BNSTpr→VMHvl and BNSTpr→MPOA projections control sex representations in these two hypothalamic nuclei. Nevertheless, we have included a caveat to state that we cannot formally exclude indirect influences of BNSTpr silencing on the altered representations in VMHvl or MPOA (**pg. 11)**.

Comment 4. "The above concern is also applied to behavior. When the BNST-Esr to -VMH-Esr or to MPOA-Esr is specifically silenced, what is the effect on behavior?"

Response: We have now performed these experiments, as suggested by the reviewer. We have silenced BNSTpr^{Esr1} terminals in VMHvl and MPOA using

Halorhodopsin, during resident intruder assays. Silencing of the BNSTpr^{Esr1} → VMHvl projections strongly inhibited ongoing aggression towards males, while causing a modest reduction in the transition from approach/sniff to mounting towards females; Silencing of the BNSTpr^{Esr1} → MPOA projections strongly inhibited the transition from approach/sniff to mounting towards females, while having little effect on male-male aggression. These results phenocopy the effects of silencing BNSTpr cell bodies (**Fig. 1**) and therefore cannot be ascribed to paradoxical effects of eNpHR3.0 at nerve terminals. We have now included these results in the revised manuscript **pg. 3** and in **ED Figure 2.**

Comment 5. "The rationale to focus on the Esr population is not clear. Are the ESR negative subset of AB neurons not responding to male and female cues or projecting to the MPOA and VMH? If not, what is their role in the model?"

Response: We focused on the Esr1$^+$ populations in all 3 structures because 1) these neurons have been functionally implicated in sniffing, mounting and attack by perturbation experiments in VMHvl and MPOA; 2) it allowed us to silence Esr1$^+$ neurons in BNSTpr while simultaneously imaging Esr1$^+$ neurons in VMHvl and MPOA, using the same Esr1-Cre driver, simplifying the genetics; Aromatase-Cre is not useful for that purpose.

In response to the reviewer's question, AB neurons constitute a relatively minor population in BNSTpr. Aromatase marks a subset of Esr1$^+$ neurons, and the Esr1 negative subset of AB neurons is even smaller. [REDACTED] The curiosity-driven question of determining the role of the Esr1-negative subset of AB neurons would require complex, expensive and time-consuming genetic intersectional strategies, and is tangential to the central point of this paper.

Comment 6. "Many of the BNST-Esr+ neurons are not active in the presence of either a male or female (Figure 2). Are these neurons also projecting to the MPOA or VMH? The use of mating and aggression as a behavioral proxy for sex identification is not very granular. Maybe the BNST contributes to other aspects of social motivation that result in the

phenotypes in figure 1. *Do the authors have another measure for sex recognition that could support the functional conclusions of the Esr neurons driving sex recognition*?"

Response: In response to the reviewer's initial question, we believe that most of the BNSTpr^{Esr1} neurons that are not active in the presence of either a male or female are either interneurons, or cells that do not project to VMHvl or MPOA. To confirm this, we have performed a new imaging analysis of BNSTpr^{Esr1} → VMHvl and BNSTpr^{Esr1} → MPOA projection neurons (labeled by retrograde delivery of GCaMP), and show that most of the back-labeled BNSTpr^{Esr1} neurons respond to either male or female cues (**ED Figure 9g-l**).

To address the reviewer's second point (italics), we have performed optogenetic silencing of BNSTpr^{Esr1} neurons during male vs. female urine preference assays and (pencil cup-enclosed) male vs. female preference tests. Our results indicate that the preferences for female vs. male urine, as well as the preference for interacting with a female vs. a male restrained in a pencil cup, are lost upon BNSTpr^{Es1} silencing, consistent with results reported in Bayless et al. (2019). Surprisingly, however, we found that despite the loss of *preference* for female cues when BNSTpr is silenced, ultrasonic vocalizations (USVs) towards females or female urine but not males or male urine; Karigo et al. 2021 *Nature*) remained intact (**ED Figure 1o-r**).

This unexpected new finding has caused us to revise our original conclusion. It suggests that in sexually experienced males, activity in BNSTpr is not required to identify or recognize intruder sex, but rather to exhibit a *preference* for female cues over male cues. (By analogy, a person who once preferred apples to oranges may lose that preference, while still being able to identify apples vs. oranges apart by their smell.) The observation that sex recognition is still intact following BNSTpr silencing fits with our original observation that intruder sex can still be efficiently decoded from neuronal activity in either MPOA^{Esr1} or VMHvl^{Esr1} neurons following such silencing (Fig. 3l, 3m). These data suggest that the deficits in mounting and attack caused by silencing BNSTpr neurons cannot be explained by a failure of sex identification, [REDACTED]. Rather, when taken together with our new results from silencing BNSTpr→MPOa/VMHvl terminals during social interactions (see **ED Figure 2** and response to Comment 4, above), the data suggest that BNSTpr^{Esr1} silencing prevents mounting and attack as a consequence of its effects on neural activity and sex representations in VMHvl and MPOA. We have now

included these results in the revised manuscript on **pg. 10**, and illustrated them with a new diagram in **Fig. 5n**.

In summary, our new data argue for a view of BNSTpr function that is substantially different from that suggested [REDACTED]. We thank the reviewer for suggesting the key experiments that brought this important revision to light.

Comment 7. "Are all of the BNST-Esr neurons that project to the two studied targets inhibitory? If not, this could confound the model."

Response: We thank the reviewer for raising this important question. Recently published BNSTpr single-cell RNA sequencing data (Welch et al., 2019, Cell, 177, 1873–1887) indicate that 95% of the BNSTpr^{Esr1} neurons are inhibitory. Thus, it's reasonable to assume that the majority of the BNSTpr^{Esr1} projection neurons are inhibitory as well. To confirm this, we have performed new experiments using cre-dependent retrograde AAVs expressing mNeongreen or mScarlet in MPOA or VMHvl of VGAT-Cre mice. The results (**ED Fig. 9a-f**) indicate that most of the back labeled VGAT+ neurons are Esr1^{+}. We have now included both the citations and our retrograde tracing results in the revised manuscript on **pg. 2**.

Comment 8. "Is it possible to manipulate the sensory signals so that BNST and VMH activity can be monitored as they mis-identify the sex of the partner? A gender illusion? This would control for all other aspects of behavior and neural activity."

Response: We appreciate the utility of the experiment suggested by the reviewer, in principle. Unfortunately, there is no validated procedure for creating "gender illusions" in mice. However, we present dual-site (VMHvl and MPOA) fiber photometry data as a reviewer figure (Reviewer Figure 1) from triadic interactions with both a male and a female intruder, in which paradoxical male-directed sexual behavior (which could reflect a "gender illusion" of the sort referred to by the reviewer) was observed. These results are described in detail in our response to comment 9 below.

Comment 9. "The model (EDF8) suggests that absolute incoming sensory activity is weighed, more female activity = mating, more male activity = aggression. Can you manipulate this by allowing your subjects to interact with multiple females and a single male simultaneously? It looks like the BNST activity persists as long as a female is present (fig 2j).

Is this correct? *If the subject first interacts with females and then a male is also added subsequently, does the simultaneous presence of a male alter the balance of activity in the VMH*? (In both cases, I expect the presence of a male would promote aggression, but the representation of male and female should not change.)

Response: The exact experiment the reviewer requested is extremely difficult to perform, for reasons that will become apparent in the following description. As an approximation to this experiment, we have now included bulk calcium measurements acquired by dual fiber-photometry performed simultaneously in VMHvl and MPOA Esr1+ neurons in the same animal (described and validated in in Karigo et al., 2021, *Nature*), recorded during triadic interactions between a resident male, an intruder female and an intruder male present in the same cage **(Reviewer Fig. 1**), as the reviewer suggested. We provide examples from two such interactions. In both cases, the male resident first interacts with the female intruder and then interacts with the male intruder.

In the first, characteristic example (**Reviewer Fig. 1** upper recording), when the recorded male is sniffing or mounting the female (purple outlined box; red and green rasters, respectively), activity in MPOA (purple trace) is higher than in VMHvl (gray trace). The converse (VMHvl>MPOA activity) is observed when the resident is sniffing the male intruder (gray box; blue raster). Towards the end of the recording session, as the animal switches from mounting a female to sniffing a male, MPOA activity rapidly declines and VMHvl activity increases (red box).

These results from a triadic interaction confirm observations made on dyadic pairs published in Karigo et al. (2021), which showed that in the presence of a female intruder, MPOA>>VMHvl activity, while in the presence of a male intruder, VMHvl>MPOA activity. These opposite ratios of MPOA:VMHvl activity in bulk calcium measurements reflect the fact that in MPOA (as in BNSTpr), female-tuned neurons outnumber male tuned neurons by ~2:1, whereas the converse is true in VMHvl (Remedios et al., 2017; Karigo et al. 2021).

The lower recording in **Reviewer Fig. 1** shows a rare case of what appears to be a "gender illusion," as mentioned by the reviewer in comment 8. In this [different] triadic assay, following a sniffing/mounting interaction with the female (purple box, red and green rasters), the male exhibits paradoxical *sexual* mounting (marked by ultrasonic vocalizations, USVs; see Karigo et al. 2021) towards the intruder male (yellow raster). During this USV$^+$ male mounting behavior, MPOA activity initially *increases* relative to VMHvl activity (red box 1, dashed vertical line), the opposite to what is typically observed (see above description of upper recording). The fact that the resident male sings when mounting the male intruder confirms that he perceives the intruder as a female (Karigo et al., 2021). Following this male-directed USV$^+$ mounting bout, the animal resumes sniffing the male. During that period, MPOA

[REDACTED]

activity slowly falls while VMHvl activity increases (lower recording, red box 1, traces to right of dashed vertical line). This suggests the resident male now perceives this intruder as a male. Eventually, following another interaction with the female, during sniffing of the male (gray box), the relative level of MPOA vs. VMHvl activity flips, reflecting the typical pattern (upper panel): MPOA shows a characteristic decline in activity, while VMHvl activity rises.

The data in the second (lower) recording suggest that the initial attempted sexual mounting of the male (red box 1) may reflect a spontaneous "gender illusion," in

which the resident's brain mistakenly identified the male intruder as a female (MPOA activity>VMHvl activity). They reinforce our previous conclusion that activity in MPOA and VMHvl does not simply reflect sensory responses to male- or female-specific cues, but rather a percept of intruder sex (Karigo et al., 2021). These data further support the reviewer's inference that more "female activity" (MPOA>VMHvl) leads to mating and more "male activity" (VMHvl>MPOA) leads to aggression, consistent with the female- vs. male- tuning bias in MPOA vs. VMHvl, respectively. Unfortunately, the low frequency of these spontaneous "gender illusion" events (1 in ~20 mice tested) makes it impractical to perform single-cell calcium imaging during such events, due to the very large number of animals that would have to be implanted with miniscopes to image activity during even a single such behavioral event.

Comment 10. "Figure 4a, it appears that with CNO (BNST-) the female responding cells are largely spatially segregated from the male responding cells and about half of the male responding cells remain male responding. *Are these two populations, those that switch vs retain sex tuning without BNST input different molecular subsets of the VMHvl-ESR population*? On repeated trials, is the same neuron able to switch sometimes and remain stable other times, or are they set to be either flexible or fixed?

Response: The apparent spatial segregation of female- vs. male-responding cells to which the reviewer refers is not a consistent observation across animals. To our knowledge, there is no method that would allow us to track the neurons that switch vs. retain sex tuning following BNST silencing via microendoscopic calcium imaging in freely moving animals, and then determine the molecular profile of these neurons, as the reviewer suggested. (Such an experiment might be possible in head-fixed animals, but they will not perform the social behaviors studied here.) We look forward to the time when such technology has been developed. In answer to the last question, we did not perform repeated CNO trials in the same animals, in order to prevent experience-dependent changes in the animals that could confound the results.

Comment 11. "I cannot find CNO only controls on neural activity and behavior, and some quantification/analysis for the silencing of BNST-Esr expressing hm4di neurons."

Response: We have now reported the percentage of BNSTpr-Esr1 neurons that express hM4Di, and included CNO only controls, in **ED Figure 1c, l, m.**

Comment 12. "Fig 2: Representative images of GCaMP infections in $BNST^{Esr1}$ neurons will be useful.

Response: We have now provided a representative image of GCaMP infections in $BNST^{Esr1}$ neurons in **ED Figure 3i.**

Comment 13. "Fig 3: Representative images of DREADDs infections in BNSTEsr1 neurons + GCaMP in MPOA/VMHvl Esr1 neurons will be useful."

Response: We have added the requested images in **ED Figure 5b-o**, as well as schematic diagrams to facilitate their interpretation by the reader (**ED Fig. 5a, f, k**).

Comment 14. "Ext [Data] Fig 4: "Why are error bars missing in k?"

Response: We only imaged a single animal (the only one of 2 implanted animals that exhibited adequate GCaMP expression), to confirm that separate populations of female- and male-tuned neurons exist in the BNST aromatase+ (AB) population. Therefore no error bars were shown. This experiment is simply an existence proof to demonstrate that the difference between our conclusions and those of Bayless et al. (2019) is not due to the use of different Cre drivers.

We again thank this reviewer for their insightful and thoughtful questions, and hope that our efforts to address them have improved the paper.

Reviewer # 2

We thank the reviewer for their helpful questions and suggestions, to which we respond below.

Comment 1. "It is interesting that the VMH and MPOA still display sex-biased responses after chemogenetic inhibition of the BNST. This could be due to factors discussed, such as a role for another brain region, but also could be due to incomplete BNST inhibition in chemogenetic experiments, related to either (1) the extent of coverage of the large BNST area by AAV injections, and 2) the Cre lines used. For (1), what % of Cre-expressing BNST cells are labeled and silenced by AAV injections, and for (2) are there sex-selective BNST neurons that do not express aromatase and/or ESR and could be contributing to downstream representations?"

Response: These are reasonable questions, for which we thank the reviewer. To address them, we have now included a panel showing that over 90% of the Esr1$^+$ BNSTpr neurons are labeled by AAV expressing hM4D-mCherry, as the reviewer requested (**ED figure 1c**). Single cell RNA-seq data (Welch et al., 2019, Cell, 177, 1873–1887) indicate that 25% of the inhibitory population in BNSTpr is Esr1$^+$. We cannot exclude that there would be a stronger behavioral or circuit-level perturbation if additional, non-Esr1$^+$ BNSTpr neurons were silenced. That said, if silencing the Esr1$^+$ subset of BNSTpr neurons was inadequate to strongly perturb the system, then we would have expected to see NO behavioral phenotype. To the contrary, we observed very robust and highly penetrant behavioral phenotypes, including the loss of preference for female over male cues, and the inhibition of both mounting and attack **(Fig. 1f, k and ED Fig. 1b).**

Comment 2. "The authors should ensure that AAV injection in the BNST does not label neurons in the VMH or MPOA (for example by retroactive labeling of neurons providing feedback control), which could confound interpretations."

Response: The reviewer is correct to bring up this important concern. We have added a figure demonstrating the lack of cell body labeling in VMHvl and MPOA following AAV injections into BNST in **ED Figure 5a-o**.

Comment 3. "The authors use a nicely comprehensive set of stimuli in Figure 2j, and showed that BNST response variance due to intruder sex was larger than the associated behavior. It would be worth discussing these observations in the context of the lab's previous work on VMH/MPOA. Is there an increased response variance due to behavior in the VMH and MPOA, suggesting further input transformation as information moves to the hypothalamus?"

Response: The reviewer brings up a great point. There is indeed an increased fraction of variance explained by behavior in MPOA, relative to BNST, and we have now mentioned those data in the **revised manuscript pg. 9 and Figure 5g**. In VMHvl, the fraction of variance due to intruder sex is larger than that due to behavior, as described in Karigo et al. (2021). These results are now summarized in a new diagram in **Fig. 5m.**

Comment 4. The authors focus here on neuronal representations in male mice; do similar sex biases in BNST and VMH exist in female mice?

Response: This is an important question that we appreciate, but it would require repeating all of the imaging experiments in female mice, with animals analyzed at different stages during the estrus cycle. This represents a large amount of additional work that likely would raise more questions than it would answer. We therefore feel it is more appropriate for a separate, follow-up study.

Comment 5. "Both the title and last sentence of the abstract should be edited for clarity to make the manuscript more accessible to a general audience. (I would go with something like 'Transformations of sex representations in the ascending limbic system' but of course this is just a suggestion!)"

Response: We appreciate the suggestion and have modified the title along the lines of the reviewer's suggestion.

Comment 6. "For ED 4b, it seems like activity is synchronized to particular events in the time series- if true, it would be helpful to provide annotation of whether such synchronized events correspond to sniffing or other social episodes."

Response: We thank the reviewer for this suggestion, and have now added such behavioral annotation to **ED Figure 4b.**

Reviewer #3

We thank this reviewer for taking the time to read the manuscript carefully, and for their thoughtful comments.

Comment 1. "The neuroimaging-chemogenetic data set (the main part which provide novel findings), although interesting, is purely descriptive and missing any mechanistic explanation (beyond the proposed hypothesis). Namely, how does sex-biased neuronal coding in the MPOA/VMH/BNST (or the altered sex-bias in the VMH following chemogenetic silencing) encode sex-typical stimuli and control different reproductive behaviors (mating and aggression) towards males and females?"

Response: We respectfully disagree with the reviewer's characterization of our experiments as "purely descriptive." We have performed a genetically based perturbation experiment (silencing of BNSTpr) and have characterized the

phenotypes of that perturbation at both the behavioral and circuit levels, in two different downstream targets of BNSTpr, a first-in-class study. Our observations reveal an unexpected transformation performed by BNSTpr^{Esr1} neurons (inverting the ratio of female:male-specific neurons in VMHvl, relative to MPOA).

We originally provided two complementary mechanistic hypotheses to explain how the observed perturbations of neural activity could explain the observed perturbations in social behavior, following BNSTpr silencing. [REDACTED] In the interests of distinguishing these hypotheses, and thereby gaining further mechanistic insight as the reviewer requested, we have performed an extensive additional series of ChemoScope experiments to monitor activity in MPOA and VMHvl before vs. after silencing BNSTpr, during unrestrained male-male and male-female social interactions. (As the reviewer may recall, our original ChemoScope experiments were performed using dangled intruders, preventing free social interactions). In this experiment, we performed the DREADD-mediated silencing of BNSTpr^{Esr1} neurons unilaterally (rather than bilaterally as in the case of behavioral assays, e.g., Fig. 1). Because there are very few inter-hemispheric connections in this circuit, the unperturbed contralateral side of the brain is able to compensate for silencing on the ipsilateral (imaged/silenced) side, and therefore social behavior is intact. This design eliminates the confound that any observed changes in neural activity in MPOA/VMHvl are simply a reflection of changes in behavior. It also allows us to temporally correlate alterations in MPOA or VMHvl activity caused by silencing BNSTpr, with specific behavioral states or transitions made by the animal. The results of these experiments are now presented in a new **Figure 5**, as well as in **ED Figures 8** and **9.**

These new data include unexpected results that strongly favor one hypothesis over the other, and provide mechanistic insights. There are 3 basic findings we report:

1)[REDACTED] a requirement for BNSTpr in sex recognition was inferred from the observation that a *preference* for female over male cues was lost when Aromatase^{+} neurons in BNSTpr were ablated or chemogenetically silenced. We

have now replicated this result with optogenetic silencing of BNSTpr^{Esr1} neurons (**ED Fig. 8i, j, l**). However, we also performed an additional behavioral test [REDACTED] namely recording of ultrasonic vocalizations (USVs). It is well known that male mice emit USVs in response to females and female cues (e.g., urine), but not in response to male cues. To our surprise, BNSTpr silencing did not impair ultrasonic vocalizations (USVs) emitted specifically in response to female urine or caged females (**ED Fig. 8k, m**).

This important result suggests that BNSTpr^{Esr1} neuronal activity is not required for males to identify and distinguish females from males. Rather it is required for males to exhibit a *preference* for interactions with females over males. While BNSTpr may play a redundant role in some aspects of sex identification, this new result indicates that the effect of BNSTpr silencing to prevent mounting and attack *cannot* simply be a secondary effect of a failure of sex identification. Therefore, they strongly argue [REDACTED] that the loss of mounting and attack behavior reflects changes in activity within MPOA and VMHvl consequent to BNSTpr silencing, that prevent them from promoting these consummatory behaviors.

2) Our new ChemoScope experiments in freely behaving animals have yielded insights into the cellular changes responsible for these behavioral phenotypes. In control animals, largely distinct subsets of MPOA^{Esr1} neurons are activated as the animals perform sniffing vs mounting. As mounting is initiated, sniff-selective neurons exhibit decreased activity, while mount-selective neurons show increased activity (**Fig. 5a, pre-CNO; 5b, left panel**). In contrast, when BNSTpr^{Esr1} neurons are silenced, this switch in active neural subpopulations is suppressed: sniff-selective cells continue to be active during mounting (**Fig. 5e, f**), while mount-selective neurons show reduced activity (**Fig. 5c, d**).

These data suggest a mechanistic explanation for the failure of males to transition from sniffing to mounting when BNSTpr^{Esr1} neurons are bilaterally silenced: during sniffing, sniff-selective neurons exert feed-forward inhibition onto mount-selective neurons. As the animal transitions from sniffing to mounting, inhibitory input from BNSTpr^{Esr1} neurons suppresses the activity of sniff-selective neurons, causing dis-inhibition of mount-selective neurons (diagrammed in **Fig. 5n**). This model is consistent with the fact that >80% of MPOA neurons are GABAergic (Moffitt et al., 2018). It is also consistent with the observation that when BNSTpr^{Esr1} neurons are

bilaterally optogenetically silenced, the animals continue to sniff, rather than mount (**ED Fig. 1b**). While these new results do not prove that the sniff-selective MPOA neurons are the same cells as the ones proposed to inhibit the mount-selective neurons, this is the most parsimonious interpretation of the data and is in principle a testable hypothesis.

3) In VMHvl, the results are somewhat different, since there are very few sniff-selective vs. attack-selective cells in this nucleus; most cells display mixed selectivity for these two behaviors (Remedios et al., 2017; Karigo et al., 2021). Correspondingly, the transition from sniffing to attack does not involve a major change in *which* subset of neurons is active. Rather, the activity of VMHvl^{Esr1} neurons that are normally active during either the sniff or attack phases becomes significantly decreased during attack (**Fig. 5j-l, ED Fig. 8b-d**). We think that this decrease in activity, together with the reduced number of male-tuned neurons caused by the inversion of the male:female sex tuning bias, prevents activity in VMHvl from reaching a threshold necessary to transition from sniff to attack (**Fig. 5n**). This model is consistent with the observation in our original optogenetic study that the transition from sniffing to attack requires a threshold level of VMHvl^{Esr1} activation (Lee et al., 2014).

Together these 3 pieces of new data argue that the requirement for BNSTpr^{Esr1} neuronal activity in mounting and attack is not simply a secondary consequence of a failure of conspecific sex identification, as proposed by Bayless et al. (2019). Rather, they reflect a change in patterns of neural activity in MPOA and VMHvl that prevent the transition from sniffing to mounting or attack, respectively, albeit by different potential mechanisms. We have incorporated these findings into the manuscript in **pg. 10-11** and as new summary diagrams (**Fig. 5m, n)**.

A further “mechanistic explanation” of the observed phenotypes and of BNSTpr^{Esr1} function in normal animals will require analysis of specific synaptic connections between these neurons, and their direct and indirect targets in MPOA and VMHvl, as well as perturbation of those specific connections. We respectfully submit that such studies are far beyond the scope of this already data-dense paper.

Comment 2. “it is essential to confirm the findings with additional complementary manipulations such as chemogenetic or optogenetic activation of the same neuronal population.”

Response: We share the reviewer's desire to complement loss-of-function with gain-of-function experiments, in principle. However, upon further consideration we realized that such experiments would be difficult to interpret. BNSTpr^{Esr1} neurons are necessary for both male-directed and female-directed consummatory behaviors. We show that separate populations of these neurons respond to either male or female cues and likely underlie this dual behavioral requirement. We have no way at present to separately activate these two sex-tuned BNSTpr^{Esr1} sub-populations. Therefore, any gain-of-function manipulation would simultaneously activate both cell types. This is a highly un-physiological manipulation that could either promote mixed behaviors, or no change in behavior at all, due to the two effects cancelling each other out. Therefore, we feel that this would not be an informative experiment, and did not constitute an efficient utilization of the limited numbers of animals available to us during the COVID pandemic.

Comment 3. The use of restrained individuals as social stimuli (first set of experiment) is very problematic, as it may trigger a massive stress response in both mice. If the authors wish to examine the response to a conspecific mouse without enabling attack or mount responses, they can present the conspecific separated by a perforated barrier or use a stimuli such as urine or soiled bedding.

Response: We take the reviewer's point, and have repeated our experiments using both urine and conspecifics separated by a perforated barrier as stimuli (**ED Fig. 8i-m**). We have also provided a figure for the reviewer showing that responses in BNSTpr^{Esr1} neurons to mouse urine are similar to those obtained using dangled male or female mice as stimuli (see **Reviewer Figure 2**). Given these data, stress responses are not likely a confound.

[REDACTED

Comment 4. The link between the behavioral and neural effects of the chemogenetic silencing is poorly explained, thus it is not clear what is the biological/functional significance of the neural findings.

Response: We refer the reviewer to our detailed response to their Comment #1 and the new data discussed therein. Our new data provide testable explanations for how the neural effects of chemogenetic silencing lead to the observed behavioral effects. This is now discussed in the revision on **pg. 10-11.**

Comment 5. "did the authors notice any sex-reversed behaviors during the silencing period (i.e. attack of subject females or sexual behaviors towards males)? These behaviors should be quantified for pre-CNO and CNO segments."

Response: We did not observe any sex-reversed behaviors during silencing. We have added panels in **ED Figure 1j, k** to demonstrate this.

Comment 6. "Also, is the silencing effect reversible? What would happen in optogenetic / chemogenetic activation of these neurons?"

Response: The effect of silencing is indeed reversible (shown in **Figure 1**, post-light period). Regarding optogenetic or chemogenetic activation, we refer the reviewer to our response to their Comment #2.

Comment 7. "In order to support their sex discrimination claim, the authors need to conduct a separate discrimination assay, during BNSTprEsr1 silencing."

Response: We thank the reviewer for this suggestion. As mentioned in response to the reviewer's comment #1, we have now performed optogenetic silencing of BNSTpr during both a urine preference test and a "pencil cup" test, wherein a male and a female intruder were placed inside the male resident's cage and separated by a perforated barrier. In parallel, we have performed recordings of USVs. To reiterate, these results reveal that while silencing eliminates the *preference* for female over male cues [REDACTED] it does *not* eliminate the animal's ability to identify and distinguish males vs. females (**ED Figure 8i-m**). Their significance is discussed in our response to the reviewer's Comment #1.

Comment 8. Do the response characteristics of specific neurons remain stable for days/weeks?

> Response: We have added a panel for the responses of MPOA mount preferring neurons tracked over 3 days in **Figure 5a**, and a panel for the responses of VMHvl attack preferring neurons tracked over 3 days in **Figure 5h**. The yield of tracked cells across weeks as detected by miniscope imaging and signal extraction is likely to be too low to be meaningful.

We hope that with the addition of these extensive new data, and the mechanistic insights that they provide, this manuscript will now be acceptable for publication in *Nature*.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

I have read the revision of this manuscript several times and even though there is a lot of experimental effort I don't agree that there is substantial insight about either the brain or behavior that is new, clear, and impactful. The stated goal of the research in the abstract is "to understand how sex is encoded at the single cell level in the BNST and how this activity influences the coding of intruder sex and behavior in the MPOA and VMHvl". The data clearly shows that BNST activity regulates the number of neurons that are preferentially tuned to male cues in the male brain in the VMH and MPOA. How to interpret in a manner that is substantially different from what has previously been published is not clear. New experiments, showing that in the absence of BNSTesr activity males appropriately USV call to female urine but not male urine, confirm that the original conclusion of these neurons (to encode sex identity) was incorrect, the animals are able to identify the sex of others with or without BNSTesr neurons. With this new understanding, the paper as it is now written is quite misleading. The intentions of the research (quoted above) and the conclusions of the first several figures are framed in a manner so that a reasonable person would conclude that these neurons do promote sex identity. In fact, the 'twist' is buried in EDF8. Furthermore, other labs have indicated the importance of the PFC to represent conspecific sex and sex preference (Kingsbury, Neuron 2020). Now that the conclusions have changed, the current state of the field and how the BNSTesr fits with known PFC data need to be experimentally evaluated or discussed. Concluding that the BNSTesr neurons are not required to encode sex identity but preference hinges on a preference experiment fashioned after Kingsbury 2020 – and again buried in EFD8. This preference data is not convincing. The variance is large, controls are lacking, and there is no statistical analysis (it does not appear significant). I am not convinced that the BNSTesr neurons drive sex preference.

The strength of this work is the precision to observe the dynamics and tuning of individual neurons during a variety of behaviors and a limited type of circuit manipulations. The observations are quite interesting to specialists (eg: there is a substantial population of neurons that are co-tuned to male and female cues – what is their function?; how does removing the inhibitory input from the BNST alter the tuning of VMH or MPOA neurons?) but these questions are very 'inside baseball' type details of interest to wonky experts. Overall, gaining single cell resolution only adds minor details, the experimental effects are subtle, the observations are still largely correlative, and the conclusions are not clear to me which makes me think they are not suitable for a general audience.

Figure 3-4 should additionally use saline control not just 'pre-CNO' to control for the effects of injection on behavior.

Referee #2 (Remarks to the Author):

The revision is improved with substantial new data, and results obtained by combining in vivo imaging with chemogenetic input manipulation are impressive. I am satisfied with the authors' response, and have no additional comments.

Referee #3 (Remarks to the Author):

Yang & Anderson have made considerable revisions to their manuscript, adding important experiments that resolve my previous concerns. Specifically, the authors have added experiments demonstrating the mechanisms by which the transition in sex bias of BNSTpr neurons affect social behaviors. In addition, the authors have added complementary experiments to the assays using restraint-intruder stimuli, where the male/female cues are presented as urine or as unrestrained mice behind a perforated divider.

I do have some minor remarks, that should be incorporated in the final version:

1. The authors have added a summarizing model (fig. 5n), however it describes only part of their results. I suggest expanding the model to include all of the dataset (or at least most of it), including the role of BNSTpr neurons and its projection to distinguish between the sexes.

2. The manuscript is focused entirely on males, however some discussion, or even an hypothesis, can be postulated as to the effects of similar manipulations in females.

Author Rebuttals to First Revision:

Nature 2021-03-04882 Yang et al. Point-by-point response

Reviewer #1

Comment 1. One point raised by [Reviewer #1] refers to the role of $BNST^{Esr}$ neurons in sex discrimination. “The authors have revised their conclusions based on new experiments of ED Fig. 8, which suggest that these neurons are instead required for the utilization of sex information to transition from appetitive to consummatory social behaviors.” It was recommended that we present those data earlier in the paper to avoid misleading the reader.

Response: We have moved the odor preference and USV tests shown in **Extended Data Fig. 8** into **Extended Data Fig. 1**. This should help clarify the paper’s framework, make the reader aware of all related behavioral observations relevant to the functional roles of these neurons concurrently, and ensure that the paper does not have a perceived 'twist' at the end.

Comment 2. “Furthermore, other labs have indicated the importance of the PFC to represent conspecific sex and sex preference (Kingsbury, Neuron 2020). Now that the conclusions have changed, the current state of the field and how the $BNST^{Esr}$ fits with known PFC data need to be experimentally evaluated or discussed.”

Response: We have now included a mention of Kingsburgy et al. (2020) Neuron and PFC and sex preference in the Discussion. (**pg. 11-12, lines 244-250**)

Comment 3. “This preference data is not convincing. The variance is large, controls are lacking, and there is no statistical analysis (it does not appear significant).”

Response: We have now performed two-way ANOVA on the dataset and now demonstrate that the preference data is statistically significant (**ED. Fig. 1r, w**).

Comment 4. “Figure 3-4 should additionally use saline control not just ‘pre-CNO’ to control for the effects of injection on behavior.”

Response: We have now clarified in the main text, figure legends and the Methods section that “pre-CNO” is where mice were injected with saline instead of CNO.

Reviewer #3

Comment 1. "The authors have added a summarizing model (fig. 5n), however it describes only part of their results. I suggest expanding the model to include all of the dataset (or at least most of it), including the role of BNSTpr neurons and its projection to distinguish between the sexes."

Response: We have re-drawn the model to emphasize the major conclusions of the paper. We feel that including further detail would complicate it and distract from its central message.

Comment 2. "The manuscript is focused entirely on males, however some discussion, or even a hypothesis, can be postulated as to the effects of similar manipulations in females."

Response: We have now included some discussion on sex representations in females. (**pg. 11-12, lines 244-250**)