JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles | Behavioral/Cognitive

Levodopa boosts hippocampal activity and strengthens generalizable rule representations in prefrontal cortex during the inference of meaning

https://doi.org/10.1523/JNEUROSCI.1263-25.2026

Received: 27 June 2025
Revised: 9 September 2026
Accepted: 16 September 2026

Copyright © 2026 Mainetto et al.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

This Early Release article has been peer reviewed and accepted, but has not been through the composition and copyediting processes.The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Levodopa boosts hippocampal activity and strengthens generalizable rule representations in prefrontal cortex during the inference of meaning

Elena Mainetto[1]*, Xiaochen Y. Zheng[1,4], Upasana Shah[1,2], Katarína Labancová[5], Robbert-Jan Verkes[2], Mona M. Garvert[3], Hanneke E. M. den Ouden[1], Roshan Cools[1,2]

[1] Radboud University, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, 6525 EN, Nijmegen, the Netherlands

[2] Radboud University Medical Center, Department of Psychiatry, 6525 GA, Nijmegen, the Netherlands

[3] Julius-Maximilians-Universität Würzburg, Faculty for Human Sciences, Marcusstraße 9-11, 97070, Würzburg, Germany

[4] Cognitive Psychology Unit, Institute of Psychology, Leiden University, Leiden, The Netherlands

[5] Department of Basic Neurosciences, University of Geneva, Switzerland

* Corresponding author: elena.mainetto@donders.ru.nl

JNeurosci Accepted Manuscript

Abstract

Dopamine has long been implicated in a variety of higher-order cognitive functions, including language processing. However, the neural processes and representations by which dopamine influences language processing remain unclear. Here we show that administration of levodopa to young healthy volunteers (n = 70, 47 females) boosts hippocampal processing and neural representations of generalizable abstract rules for successful inference of novel word meaning. We used a novel experimental paradigm that required participants to build new artificial word meanings by inferring and applying abstract affixation rules. A repetition suppression procedure was combined with multivariate representational similarity analyses to isolate the neural locus of both the abstract rules required for the compositional inference process, as well as the novel word meaning representations that were composed based on these abstract rules. Levodopa modulated not only inference-related signals in the hippocampal formation, but also abstract rule representations in the medial frontal cortex. These results support the hypothesis that dopamine boosts hippocampal activity and strengthens prefrontal representations of abstract rules when building novel meaning.

Significance statement

Dopamine has long been implicated in a variety of higher-order cognitive functions, including language processing. However, the neural processes by which dopamine boosts language processing have remained unclear. Here we show that administration of levodopa to young healthy volunteers enhances hippocampal activity and strengthens prefrontal representations of generalizable abstract rules when building novel word meanings.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

This research was supported by the Language in Interaction consortium (Gravitation Grant 024.001.006 funded by the Dutch Research Council). RC is supported by ERC Advanced grant CHEMCONTROL #101054532. The authors would like to thank Anna Aumeistere, Marwan Engels, Maartje Graauwmans and Annkathrin Böke for their help during the data acquisition.

JNeurosci Accepted Manuscript

Introduction

Dopamine has long been implicated in a wide variety of high-level cognitive functions, including working memory, rule learning and cognitive flexibility (Cools et al., 2002; Deserno et al., 2021; Valentin et al., 2016; Wunderlich et al., 2012), and drugs that increase dopamine are well known to have cognitive enhancing effects (Berridge & Devilbiss, 2011; Cools & D'Esposito, 2011; Husain & Mehta, 2011; Sahakian et al., 2015). Such enhancing effects are often attributed to increases in dopamine receptor stimulation in the prefrontal cortex (PFC), which have been shown to boost the gain of abstract rule representations in working memory (Braver & Cohen, 1999; Cools, 2016; Durstewitz et al., 2000; Ott & Nieder, 2019; Servan-Schreiber et al., 1990).

A separate literature has shown that dopamine can also enhance language processing (Copland et al., 2009; Kischka et al., 1996; Tan & Hagoort, 2020). However, the neural processes and representations that underlie these effects on language processing are unclear. The present study begins to elucidate these by focusing on the contribution of dopamine-related strengthening of generalizable, abstract rule representations to a particularly important feature of language processing: the inference of meaning. To study this contribution, we scanned a group of healthy participants after intake of placebo or the precursor of dopamine, levodopa, while they completed a task that enabled assessment of novel word meaning inference.

The inference of meaning relies on the generalization of abstract rules that define the relational structure between different linguistic elements that together constitute a word, a sentence, or a story (Frankland & Greene, 2020a). For example, we can infer the meaning of a word we have never previously heard, like 'un-google-able', because we can generalize and apply the abstract affix rule that the placement of 'un' before a word negates its meaning (Tamminen et al., 2015). Outside the domain of language, the ability to construct novel concepts, plans, memories or decisions based on generalizing abstract relational structures is well accepted to implicate the (heavily dopamine-innervated) PFC (Miller & Cohen, 2001; O'Reilly, 2006; Russin et al., 2020). Specifically, studies in the domain of relational memory and goal-directed behaviour have highlighted a key role for circuits connecting the (medial) PFC with structures in the medial temporal lobe, including the hippocampus and entorhinal cortex, in the generalization of relational structures for novel inference (Barron et al., 2013; Elston & Wallis, 2025; Knudsen & Wallis, 2021; Kurth-Nelson et al., 2023; Schwartenbeck et al., 2023). Accordingly, we hypothesized that abstract rule-based inference of novel word meaning involves dopaminergic modulation of regions in the 'relational memory' network (including the hippocampal formation and the medial PFC), as well as the language network and the striatum (where dopamine receptors are most abundant) (Forkel & Hagoort, 2024; Zheng et al., 2025). In keeping with the proposed core role of fronto-hippocampal circuitry in the dopaminergic modulation of meaning inference, accumulating evidence implicates interactions between the dopaminergic midbrain and the hippocampal formation in the composition of novel concepts (Barron et al., 2020; Kahnt & Tobler, 2016; Shohamy & Wagner, 2008).

To test this hypothesis, we used an experimental procedure previously established to enable detection of abstract rule representations for word meaning inference in lateral and medial PFC (Zheng, et al., 2025; Zheng, et al., 2024). After oral ingestion of either 150mg of levodopa or a placebo, participants generalized abstract affixation rules to infer new artificial compositional words on the fly. By combining fMRI repetition suppression and representational similarity analyses, we revealed how abstract rules and word meanings are represented and inferred as a function of levodopa. In line with our hypothesis, administration of levodopa modulated the hippocampal formation during the compositional inference process. Critically, levodopa also increased the strength of the abstract rule representations in the superior medial

Neuroscience Accepted Manuscript

PFC. This observation suggests that levodopa boosts neural activity related to hippocampal inferential processing and prefrontal representations of abstract rules when building novel meaning.

Materials and Methods

Participants

We recruited 70 healthy native Dutch speakers (47 women and 23 men, mean age = 25.4 years, SD = 6.8) to participate in the study. All participants provided written informed consent before taking part and were reimbursed 100 euros for their time. Participants were recruited via the Radboud University electronic database for research participants and via flyers around Nijmegen. The study was part of a bigger protocol that included additional tasks reported in a separate paper (*Figure S1*). The study was approved by the local ethics committee (CMO 2022-13817 Oost-Nederland) and conducted in accordance with the Declaration of Helsinki.

The study was preregistered at https://aspredicted.org/hy838.pdf. The sample size was based on a G-power analysis that indicated that 90 participants would be required to detect a medium between-group (independent t-test) effect size (d = .60), given a power of 0.80 and an alpha of 0.05 (tail(s)=2, allocation ratio N1/N2=1). We reasoned this would be a conservative approach, given that previous studies of levodopa's effects on higher cognitive functions had reported a medium or larger effect size (Copland et al., 2003; Wunderlich et al., 2012). Due to the expiration of the research grant and employment contracts, the study had to be closed prior to reaching the pre-registered number. Based on the final sample of 70 participants (levodopa n = 34; placebo n = 36), the between-subject effect sizes we can detect need to be larger than .68 (tail(s)=2, with an independent t-test, 0.8 power, 0.05 alpha).

Participants were screened for reported history of neurological, psychiatric or any other current medical conditions, in addition to being right-handed, Dutch native speakers and between 18 and 45 years old. A list of all the inclusion and exclusion criteria is presented as *Text S1*. Furthermore, the two treatment groups (placebo and levodopa) did not differ significantly in terms of age, body mass index, verbal IQ, anxiety, impulsivity, burnout and depression (**Table 1**). Surprisingly, the levodopa group exhibited increased ratings on the reward sensitivity questionnaire (Behavioural activation scale, Carver & White 1994), which was administered prior to drug administration. However, this difference in reward sensitivity between treatment groups did not reach significance after correction for multiple comparison. Moreover, control analyses revealed the effects of interest reported below to remain significant after controlling for individual differences in reward sensitivity (*Figure S2*).

Of the included 70 participants, six failed to meet the task performance criterion (described in *Behavioural data analysis*) and one was excluded due to motion artifacts in the fMRI data. This resulted in a total of 63 participants of whom 29 were administered levodopa and 34 placebo. Eight participants were further excluded from the fMRI multivariate pattern analysis (described in *fMRI multivariate analysis*), due to an error in the randomization script (explained in *fMRI univariate analysis*). This specific analysis was thus performed on 55 participants, of which 25 belonged to the levodopa group and 30 to the placebo group.

Study procedure

Journal Accepted Manuscript

The study employed a double-blind, between-subject design. A detailed description of all the measurements collected is presented in *Figure S1*. All participants attended two sessions approximately 10 days apart: a screening session and a pharma(cological) session in which they took the drug or placebo (**Figure 1a**).

The experimental paradigm of interest consisted of three phases: 1) learning phase; 2) inference phase; 3) post-test phase. In the screening session, participants completed only the first phase of the task. In the pharma session, participants completed all three phases of the task.

In the learning phase, participants were shown compositional pseudo-words. These pseudo-words consisted of a known (non-artificial) stem and an artificial affix (e.g., "goodkla") and they were shown together with their meaning (e.g., "bad") (**Figure 1b**). This phase allowed participants to infer generalizable affix rules for composing artificial word meaning based on exemplars. In the inference phase, participants applied these rules to novel pseudo-words (e.g., "whitekla") to infer their meaning on the fly (e.g., "black") (**Figure 1c**). This phase was completed during fMRI and allowed us to assess the inferential processes and the representations of word meanings and affix rules. In the post-test phase, participants had to write down the meaning of the pseudo-words seen in the inference phase. This allowed us to quantify the degree to which word meaning inference was successful (**Figure 1d**).

During the screening session, participants were screened for inclusion and exclusion criteria. They also performed in the first learning phase of the task. Furthermore, we collected a baseline measure of working memory capacity using the (forward and backward) digit span (Groth-Marnat, 1997), to assess whether the levodopa and placebo group were matched in terms of working memory capacity. There was no significant difference between the two treatment groups (*Text S4*).

Between the screening session and pharma session, participants completed a set of questionnaires at home for demographic information. This information confirmed the absence of significant differences between treatment groups such as age, body mass index, verbal IQ, anxiety, impulsivity, burnout and depressive symptoms (*Table 1*).

In the pharma session, participants first performed a shorter version of the same learning phase to refresh their memory of the learned pseudo-words. Subsequently, they ingested the drug: either levodopa (150mg levodopa /37.5mg benserazide; Madopar, Roche) or placebo. Levodopa is a dopamine precursor known to pass the blood-brain barrier and to increase dopamine brain levels. After approximately 28 minutes (SD = 7) from drug intake, participants completed the inference phase while undergoing fMRI. Drug administration timing was optimized for the peak drug effect during this phase (after about 30 minutes up to two hours) (Iwaki et al., 2015). To standardize drug absorption across individuals, participants were instructed not to eat for up to two hours prior to arrival.

Right after exiting the scanner, participants performed a second digit span to assess drug effects on working memory capacity. The list of numbers during this digit span differed from the one administered in the screening session. There were no significant drug effects (*Text S4*). At the end of the pharma session, we tested participants' knowledge of the meaning of the novel pseudo-words with the post-test phase.

Across the pharma session, we monitored medical symptoms (Määttä et al., 2021) and subjective states (Bond & Lader, 1974) at different time points. There were no significant effects of treatment group on any of the medical symptoms and mood measures ratings (*Text S2*).

Across the pharma session, we further assessed arousal and motor invigoration using a finger tapping task. The task consisted of pressing a button as quickly as possible with the right index finger after hearing a tone (1000 Hz) (Knecht et al., 2004). We did not detect any difference in reaction times between treatment groups (*Text S5*).

Neuroscience Accepted Manuscript

Finally, to validate and monitor the levodopa manipulation, we collected physiological measures of diastolic and systolic blood pressure and heart rate at four time points across the pharma session (*Text S3*).

Experimental material

The task paradigm and stimulus materials were identical to those used in Zheng et al. (2025), to allow for comparison across studies. A complete list of the stimuli is presented in *Table S1* and *Table S2*. While the materials are here illustrated using English examples, the actual materials were in Dutch

In the learning phase, 30 pseudo-words were presented, consisting of a known stem and an unknown affix (e.g., “good-kla”, “kla-dog”). Critically, the affix altered the meaning of the stem depending on its position, according to an abstract rule (e.g., “-kla” as a suffix meant “the opposite”, whereas “kla-” as a prefix meant “young version”). For the inference phase, a separate set of 30 pseudo-words (i.e., not presented in the first phase) was presented.

Overall, six abstract rules were selected. These rules were assigned to three affix forms that were presented either as prefixes or as suffixes, so that the rule associated with the same affix form depended on its position as a prefix or suffix (e.g., “-kla” as a suffix meant “the opposite”, whereas “kla-” as a prefix meant “young version”). To avoid fMRI effects being confounded by visual input, the rule-affix form assignment was counterbalanced across participants. All the phases of the task were presented using the software Presentation (Version 20.2, Neurobehavioural System Inc, Berkeley, U.S.).

Learning phase

The learning phase required participants to learn the meaning of novel compositional pseudo-words and was set up so that they could infer the rules (incidentally, not by instruction). During this first phase, participants were shown the pseudo-words and the meaning of these words together with a sentence example for a maximum of 15 seconds (learning block) (e.g., sentences were shown “goodkla” means “bad” and “kladog” means “puppy”) (**Figure 1b**). After passively viewing all the pseudo-words and the associated example sentences, their understanding and memory of these words were probed (memory block). Here, they were shown a synonym meaning with four pseudo-words and were asked to choose the matched pseudo-word. Importantly, the meaning of the pseudo-words depended on the position of the affixes (abstract rules). Participants were never instructed about these rules, but the rules were incidentally learned and generalized.

The learning and memory blocks in this first phase of the task consisted of thirty pseudo-words (5 pseudo-words for each of the 6 rules, presented in pseudo-randomized order). Pairs of learning and memory blocks were repeated four times in the screening session and twice in the pharma session.

Inference phase

After learning the pseudo-words, participants ingested the drug. Next, they performed the inference phase in the scanner. The fact that the drug was administered after the learning

JNeurosci Accepted Manuscript

phase means that any drug effects cannot reflect modulation of rule inference, but must reflect modulation of the application of rule-based inference of novel word meaning.

During this inference phase, participants were shown novel pseudo-words ("primes") that they had never seen in the learning phase. This phase therefore required inference on the fly. Crucially, primes were designed to either create conflict or not based on the application of the abstract rules to infer meaning depending on two conditions: 1) a congruent condition and 2) an incongruent condition. In the congruent condition, the prime pseudo-word was congruent with the rule (e.g., for "white-kla", "-kla" as suffix means "the opposite", and the opposite of white is "black" and congruent with the rule). In the incongruent condition, the prime pseudo-word was not congruent with the rule (e.g., for "kla-white", "kla-" as a prefix means "young version", and "young version of white" is meaningless and incongruent with the rule).

Primes were always followed by "targets", which were words with a known meaning. Targets were always synonyms corresponding to the meaning of the congruent version of the prime (e.g., the target "black" was presented after "white-kla" as well as "kla-white"). Across the inference phase, participants were instructed to imagine the meaning of both the prime and target words when these were presented on the screen.

The inference phase also included a control condition. In this control condition, the stems of the primes were combined with alternative affixes, so that regardless of the position of affix (i.e., the application of the rule), the meaning of the prime was most likely uninterpretable and not a synonym of the target (e.g., ran-white = "the colour of" white / white-ran = "the extreme version of"; both are uninterpretable and do not match the meaning of the target word "black"). The control condition was relevant for: 1) identifying participants who failed to learn the abstract rules, because the meaning of this condition pseudo-words is uninterpretable regardless of the affix position (see also *Behavioural data analysis*; and 2) balancing the frequency of the affix positions across the inference phase. The interpretability of the pseudo-words in the three conditions was validated in a series of experiments in an early study (Zheng, et al., 2024).

Overall, the inference phase entailed 2 blocks, each with 120 trials of prime-target pairs with the conditions equally balanced across each block. Specifically, a block consisted of 30 congruent trials, 30 incongruent trials, 60 control trials for affix position (30 for pre/post affix position, like "white-ran" or "ran-white"). Each prime was presented for 1.5 seconds with a jittered inter-trial interval displaying "***". A prime was followed by a target shown again for 1.5 seconds, followed by a jittered inter-trial interval displaying "+". The jittered intervals were generated from a truncated exponential distribution with a mean of 2.5 seconds. Importantly, all the pairs of prime and target were presented in a pseudorandom order by: 1) repeating the same affix form or affix position on maximally three consecutive trials; 2) repeating the same condition on maximally three consecutive trials; 3) repeating the same stem at least five trials apart.

To ensure that participants paid attention and made use of the abstract rule to infer the new meaning, a probe question was presented after the presentation of the target on 10% of the trials (i.e., 24 probe trials per participant). These probe questions required participants to indicate whether the meaning of the prime pseudo-word matched that of the target word. They responded by pressing the left ("yes") or the right ("no") button on the button box. The probe questions stayed on the screen for a maximum of 10 seconds or till participants responded.

<u>Post-test phase</u>

JNeurosci Accepted Manuscript

At the end of the pharma session, participants reported for every pseudo-word they saw in the inference phase whether it was meaningful and, if so, what it meant (comprising all the 120 words encountered in the congruent, incongruent and control condition). We further included the 30 pseudo-words encountered in the first learning phase as a sanity check of participants' memory of the instructed pseudo-words (instructed condition). At the end of the post-test, participants were explicitly asked to write down the meaning of the three sets of affixes.

Behavioural data analysis

The behavioural data analysis pipeline in R (R core team, version 4.3.2) followed the one used by Zheng and colleagues (2025). In both the inference and post-test phase, we performed generalized mixed models ("lme4", Bates et al. 2015) to investigate condition and treatment group effects. In these models, the optimizer was set to "bobyqa" and family to "binomial". Significance of the main effects were determined using the "afex" package (Singmann et al., 2016), while significance within a factor was assessed using the Tukey's test with "lsmeans" (Lenth, 2016). Linear regressions between behavioural performance and fMRI BOLD responses in following analyses were performed using the R package "stats".

Learning phase

To assess whether participants learned the meaning of the pseudo-words in the learning phase, we investigated participants' accuracy in the pharma session. In this session, accuracy in the last memory block of the learning phase was evaluated with a one-sided t-test for above chance level accuracy (25%).

Inference phase

The ability to infer novel word meaning on the fly was assessed by investigating the effect of congruence condition (congruent, incongruent, control) on the probability of 'meaning match' responses on the 10% probe trials. These match response rates were submitted to a generalized mixed effects model to test the main effect of condition (congruent vs. incongruent vs. control) and its interaction with treatment group (placebo vs. levodopa). Participants and items (grouped by the stem of the pseudo-words) were included as random effects. Treatment group was contrast coded relative to the grand mean, while condition was dummy coded (with the congruent condition as baseline). It should be noted that this phase design with only 8 probe trials per condition was optimized for detecting neural effects, but not behavioural effects.

Post-test phase

Yes/no responses to the question "Is the [pseudo-word] meaningful?" and free responses to the question "What does [the pseudo-word] mean?" were analysed to substantiate participants' ability to infer the meaning of the novel pseudo-words (presented during the inference phase). To this end, we first transcribed the written free responses as follows: 1) "Matching the synonym"; 2) "Meaningless"; 3) "Creative", if the answer was unexpected (e.g., when one consider a pseudo-word "human-kla", the opposite of human, to be "animal"); 4) "Unexpected but incorrect", if the meaning of the affix form were confused (e.g. if "warm-ran" was mistaken as "warm-kla" and the reported meaning was "cold", the

JNeurosci Accepted Manuscript

opposite of warm). After transcription, we excluded the unexpected free responses (3 and 4), which concerned an average across participants of 6.72% of the trials. Finally, in order to calculate the mean proportion of 'yes, meaningful' responses and investigate treatment group effects on the ability to correctly infer word meaning, we coded the "Matching the synonym" and "Meaningless" as 1 and 0, respectively. The pseudo-words of the instructed condition, instead, were transcribed as either "Matching the synonym (presented during learning phase)" and coded as 1, or "Wrong" and coded as 0.

Based on the post-test, a total of 6 participants were excluded from all analyses, because they failed to learn and/or apply the affix rules to correctly infer the meaning of the pseudo-words (either during the learning phase or during the key inference phase). Participants were excluded if they met any of three criteria: (1) they judged more than 50% of control pseudo-words as meaningful (n = 1); (2) they judged fewer than 50% of instructed pseudo-words as meaningful (n = 2); or (3) they judged fewer than 50% of congruent pseudo-words as meaningful and failed to report at least one correct rule in two of the three affixes (e.g., participants did not infer the meaning of the prefix "kla-" and suffix "-run" for the affixes "kla" and "run" respectively) (n = 3).

Both the yes/no responses and the transcribed free responses were submitted to generalized mixed effects models, with condition (congruent vs. incongruent vs. control) and treatment group (placebo vs. levodopa) as (interacting) independent variables. Participants and the items (grouped by the stem of the pseudo-words) were included as random effects, with condition as a random slope for participants. Finally, condition was set to dummy coding (with the congruent condition as baseline), while treatment group was set to contrast coding relative to the grand mean.

fMRI acquisition and preprocessing

MRI data were acquired using a 32-channel head coil on a 3T MAGNETOM Prisma[Fit] MR scanner (Siemens AG, Healthcare Sector, Erlangen, Germany). A single Echo-Planar Imaging (EPI) sequence with a multi-band acceleration factor of 4 was used to acquire 68 slices per functional volume (voxel size: 2 x 2 x 2 mm; Repetition Time: 1500ms; Echo Time: 37ms; Flip Angle: 75˚, Eco spacing: 68ms). The first six volumes of each block were discarded to allow for scanner equilibration. After the experimental part, a high-resolution structural image with 1x1x1 mm resolution was acquired (MP-RAGE; Repetition Time: 2300ms; Echo Time: 3.03ms; Flip angle: 8˚).

The visual stimuli during the inference phase were projected onto a screen via a computer monitor. To respond to the probe trials, participants were given an MRI-compatible button box positioned on their belly. They were instructed to respond with the index and middle finger of their right hand.

MRI data quality control was performed using MRIQC 22.0.6. (Esteban et al., 2017). Means of framewise displacement (both in mm and as a percentage of timepoints), temporal SNR, and DVAR for functional images were computed per participant based on the image quality metrics. Blocks with any of these values larger than 3 SD from the group mean were excluded (or smaller than 3 SD for temporal SNR).

MRI anatomical and functional data were preprocessed using fMRIprep 22.0.1 (Esteban et al., 2019), following the steps specified in *Text S6*. The subsequent functional data analyses (including spatial smoothing with a 6 mm full width half maximum three-dimensional Gaussian kernel) were performed using SPM12 (Statistical parametric mapping, http://www.fil.ion.ucl.ac.uk/spm) on MATLAB 2022b.

JNeurosci Accepted Manuscript

fMRI univariate analyses

First level

To test univariate fMRI effects, we implemented an event-related generalized linear model (GLM) including 10 explanatory variables. Crucially, we modelled the BOLD responses to both prime and target event onsets separately depending on the experimental conditions: congruent, incongruent and control (separately for pre/post affix position, i.e. separately for "kla-" and "-kla"). Two of the 10 explanatory variables were of no interest: the onset of the probe trials and the onset of the motor responses to those probe trials. Additional nuisance regressors were included in the GLM to account for motion-related artifacts obtained from fMRIprep (both rotation and translation in x, y, and z; framewise displacement; the first five noise components calculated with an anatomical component-based noise correction method, called CompCor; all cosine regressions). Finally, the two blocks of the inference phase were modelled separately within the GLM. For 12 out of the 63 participants, we only analysed one of the two blocks (of the 12 participants: four could not complete one of the blocks owing to technical errors or time constraints; for eight, there was an error in the trial presentation file which led to an exclusion of half of the trials in each of the two blocks, corresponding to a total of one block). The contrast images of the participants were then entered into a second-level random effects analysis.

Second level

We first analysed the main task effects, averaged across treatment groups, to assess their consistency with those reported in the previous fMRI study with this task (Zheng et al., 2025). These analyses aimed to isolate the neural locus of (1) the inference process itself, and (2) the novel meaning representations that were inferred based on abstract rules. Given that congruent pseudo-words were behaviourally identified as meaningful more often than incongruent pseudo-words, we concluded that participants applied the abstract rules (i.e., based on the position of the affix). Accordingly, the control condition was not included in further analyses.

To isolate neural signalling related to the inference process, we compared BOLD signal change at the time of congruent primes versus incongruent primes. To isolate neural representations of the novel word meanings, we leveraged the phenomenon of fMRI repetition suppression. This phenomenon relies on the fact that neuronal responses are suppressed upon repeated presentation of stimuli to which they are sensitive (Barron et al., 2016). Specifically, we assessed putative repetition suppression effects at the time of the outcome of the inference process (at target) by comparing BOLD responses to a target (e.g., "black") that followed a congruent prime (e.g., "white-kla", which means "black") with BOLD response to exactly the same target (e.g., "black") that followed an incongruent prime (e.g., "kla-white", which does not mean "black"). This repetition suppression effect is akin to the well-known mismatch effect due to different expectations between congruent and incongruent targets. The fact that the target presented was the same in all conditions (e.g., congruent condition: the prime "withe-kla" is followed by the target "black"; incongruent condition: the prime "kla-white" is followed by the target "black") allowed us to control for repetition suppression effect related the meaning of the prime stem (e.g. white) and the target (e.g. black).

Subsequently, the first-level contrast images of each participant were entered into a second-level two-sample t-test (random effects) to compare treatment groups (levodopa vs. placebo). We assessed treatment group effects related to the inference process (at the time

J Neurosci Accepted Manuscript

of the primes) and to the outcome of the inference process (repetition suppression effects at the time of the targets).

Regions of interest

For treatment group effects only, we conducted targeted region of interest (ROI) analyses using small volume correction (SVC) in: 1) the hippocampal formation (incl. hippocampus, entorhinal cortex, and subiculum; anatomically defined mask from the Juelich histological atlas, with a probabilistic threshold of 50%), given our a priori hypothesis about the role of the hippocampal formation and dopamine in inference; 2) the striatum (incl. the caudate nucleus and putamen; from the Harvard-Oxford cortical structural atlas), given the disproportionate abundance of dopamine receptors in the striatum. These two ROIs were in line with our pre-registration. Exploratory analyses further assessed effects within an anatomically defined mask of the left inferior frontal gyrus (left IFG; Harvard-Oxford cortical structural atlas, with a probabilistic threshold of 30%). This exploratory ROI (left IFG) analysis was motivated by the results from our previous study with this task (Zheng et al., 2025), which were obtained after preregistration. These results showed prime- and/or target-related (patterns of) activity in the left IFG (as well as the hippocampal formation and striatum).

Statistics

We evaluated the whole-brain effects as significant if they survived family-wise error (FWE) at the cluster-level of $p < .05$ (with a cluster-defining threshold of .001 uncorrected). We evaluated the results of SVC as significant if they survived family-wise error (FWE) correction at the peak-level of $p < .05$ (with a cluster-defining threshold of .001 uncorrected). The reported anatomical labels are based on the MNI space coordinates obtained with SPM12 from the Automated Anatomical Labeling (AAL) atlas.

fMRI multivariate analyses

To decode the (effects of levodopa on the) neural representations of both the abstract rules (e.g., “-kla” reverses the meaning of the stem word) and the newly inferred word meanings (“white-kla” means “black”) at the time of the prime, we adopted a multivariate RSA approach (Kriegeskorte, 2008; Nili et al., 2014). Consider the pseudo-word “white-kla”: To infer its meaning, participants would need to represent the *rule* (“-kla” means “the opposite of”). Moreover, provided successful inference, they will also represent the inferred word *meaning* (“white-kla” means “black”). RSA allows us to capture the relevant neural representations by computing the neural representational dissimilarity matrices (RDMs) based on prime-related fMRI activity for each pseudo-word, analysed through a whole-brain searchlight. We assessed whether these neural RDMs were explained by model RDMs (see below) that capture the similarities between these pseudo-words as a function of either the rule that was used to inferred them (by experimental design) or their inferred meaning (derived from a word embedding model). Similar to the univariate analysis, the multivariate analysis focused on the congruent and incongruent conditions.

Neural RDMs

To capture the multivariate activity patterns elicited by the prime pseudo-words (i.e., to compute pairwise neural similarities between the different primes), we ran two GLMs. The first

J Neurosci Accepted Manuscript

GLM (named “congruent & incongruent RDM”) modelled the prime events and contained separate onset regressors for each of the 30 pseudo-words regardless of whether they were congruent or incongruent (e.g., a single onset regressor was used for both “white-kla” and “kla-white”). The second GLM (named “congruent vs. incongruent RDM”) again modelled the prime events, but separately for the congruent and the incongruent conditions (e.g., one onset regressor was used for “white-kla” and another one for “kla-white”). Both GLMs contained one regressor for all the prime events in the control condition, one regressor for all the target events, one regressor for the probe trials, and one regressor for the button presses (the last two were regressors of no interest). All regressors were convolved with a canonical hemodynamic response function. Both GLMs included the same confound regressors as in the univariate analysis. Each block was modelled separately within the GLMs. Subsequently, we computed the neural RDMs for the 30 items at prime from the parameter estimate maps derived from the GLMs described above. Pairwise correlation distance (one minus Pearson correlation coefficient) was used as the distance metric. RSA was performed whole-brain using a searchlight approach with a spherical searchlight (radius = 7 mm, approx. 180 voxels).

Rule model RDM

We further constructed a rule model RDM that captures the representations of the affix rules, in which all pseudo-words ending with “-kla” are more similar to each other than the pseudo-words with a different affix (e.g., “kla-”, “-ran”, or “ran-”). We reasoned that for brain regions where these abstract affix rules are represented, their BOLD pattern should be best explained by the rule model. For this rule model RDM, we considered a 30*30 binary-coded distance matrix where the rules are either the same (e.g., “good-kla” and “short-kla”), or different (“good-kla” and “kla-dog”). It is worth noting that while the rule model was designed to capture abstract rule representations, it may also reflect affix-based similarity driven by the visual properties of the stimuli.

Meaning model RDM

We constructed a meaning model RDM that captures the representations of the newly inferred word meanings, in which word meanings are arranged by their semantic similarities derived from a word embedding model (e.g., “black”, derived from “whitekla”, is more similar to “dark” - derived from “lightkla” - compared with “happy” - derived from “sadkla”). We reasoned that the similarity of BOLD patterns in brain areas that represent the newly inferred word meanings (e.g., white-kla” means black) should reflect the semantic similarity of the composed words (e.g., “black”), as derived from this meaning model. The meaning model RDM was constructed using the embedding vectors for the 30 target words (e.g., “black” in “white-kla = black”) from a word embedding model showed to best fit behavioural data (Mandera et al., 2017). Word embedding provides learned representations in a vector space for text where words that have similar meanings have similar representations. We calculated the pairwise Pearson correlation distance between the embeddings of all target words and constructed a 30 *30 distance matrix for target meaning representations.

Alternative model RDMs

We further assessed two alternative models: one describing the similarity between the meanings of the stems contained in the prime pseudo-words; the second one describing the visual similarity of individual words based on pixel-wise Euclidian distance. These two alternative models could not account for the patterns of activity derived from the meaning model RDM (*Text S8*).

Neurosci Accepted Manuscript

Statistics

Firstly, within each searchlight sphere in each participant, we compared the model RDMs and the neural RDMs using Kendall's rank correlation. Both the searchlight of the neural RDMs and the comparison with the model RDMs were performed using the MATLAB-based RSA toolbox (Nili et al., 2014). For the congruent vs. incongruent neural RDM (which modelled separately the congruent and the incongruent conditions), the relative correlation maps were further subtracted to compute the difference between congruent and incongruent conditions.

Secondly, we analysed the main task effects (averaged across treatment groups) to assess their consistency with those reported in the original fMRI study with this task (Zheng et al., 2025), by submitting the resulting correlation coefficients to a one-sample t-test using SPM12 (i.e., contrasting the obtained correlation against zero).

Lastly, to assess the effects of treatment group, we submitted the correlation coefficients to independent sample t-tests. Effects were tested for statistical significance using cluster-inference with a cluster-defining threshold of $p < .001$ and whole-brain cluster-level FWE correction at $p < .05$.

Results

Novel word meaning inference was successful

We began by validating that the levodopa manipulation and the task manipulation were effective. The levodopa manipulation was assessed by investigating physiological measurements, medical symptoms and subjective states recorded at different time points during the pharmacological session (see *Text S2* and *Text S3*). The task manipulation was assessed by investigating participant performance depending on task conditions (i.e., congruent, incongruent, control) in the learning phase, the inference phase (in the scanner) and the post-test phase.

The levodopa manipulation was validated by establishing significant effects on diastolic blood pressure measurements when drug levels peaked. This was shown by a significant interaction between time points and treatment groups (mixed ANOVA: $F(3) = 3.72$, $p = .017$) and relevant significant simple main effects (right after exiting the MRI scanner: $mean_{Levodopa} = 68.6$, $SD_{Levodopa} = 7.0$; $mean_{Placebo} = 72.7$; $SD_{Placebo} = 8.2$; $t(67) = -2.23$, $p = .029$). The full report is presented in *Text S3*.

The task manipulation was validated by demonstrating that participants learned the abstract rules and applied them to infer novel word meaning. This was firstly evident from the performance on the memory blocks of the learning phase: These confirmed that participants learned the meaning of the pseudo-words (mean proportion correct = 97%, SD = 5.5, $t(69) = 71.49$, $p < .001$). Secondly, performance on the probe trials during the inference phase demonstrated that participants successfully generalized the rules and correctly inferred the meaning of the novel pseudo-words (congruent: mean = 88.5%, SD = 14.8; incongruent: mean = 28.0%, SD = 33.0; control: mean = 11.7%, SD = 15.0). Specifically, the average rate of 'meaning match' responses was higher for the congruent than incongruent pseudo-words ($Z = 14.24$, $p < .001$), for the congruent than control pseudo-words ($Z = 19.10$, $p < .001$), and for incongruent than control pseudo-words ($z = 1.60$, $p < .001$) (main effect of condition with three levels: $\chi^2(2) = 474.67$, $p < .001$) (**Figure 2a**). Thirdly, the proportion of 'yes, meaningful' responses to the 'instructed' pseudo-words in the post-test phase confirmed that participants

JNeurosci Accepted Manuscript

remembered the meaning of the pseudo-words that had been presented in the learning phase (mean instructed condition = 91.8%, SD = 9; above 50% chance level t-test: t(*62*) = 36.96, p < .001). Additionally, in this post-test phase the proportion of 'yes, meaningful' responses was significantly higher for congruent pseudo-words than for incongruent or control pseudo-words (congruent: mean = 84.1%, SD = 13.6; incongruent: mean = 18.6%, SD = 25; control: mean = 15.3%, SD = 11.4) (main effect of condition: $\chi^2(2) = 141.53$, p < .001; congruent vs. incongruent: Z = 11.16, p < .001; congruent vs. control: Z = 21.12, p < .001; incongruent vs. control: Z = 2.75, p = .017). Lastly, effects on the transcribed responses in the post-test phase similarly demonstrated that participants successfully inferred the meaning of the novel pseudo-words using the rules (congruent: mean = 83.3, SD = 14.1; incongruent: mean = 16.7, SD = 24.8; control: mean = 3.8, SD = 8.0) (main effect of condition: $\chi^2(2) = 113.14$, p < .001; congruent vs. incongruent: Z = 9.33, p < .001; congruent vs. control: Z = 16.23, p < .001; incongruent vs. control: Z = 1.53, p = .277) (**Figure 2b**).

In contrast to our hypothesis, however, there was no evidence for sensitivity of these behavioural indices to modulation by levodopa (**Figure 2**). Specifically, there was no evidence for treatment group effects on: (1) the 'meaning match' questions asked during the probe trials of the inference phase (main effect of treatment group: $\chi^2(1) = 0.04$, p = .843; condition x treatment group interaction: $\chi^2(2) = 0.49$, p = .784); (2) the post-test 'yes, meaningfulness' questions (main effect of treatment group: $\chi^2(1) = 1.96$, p = .161; condition x treatment group interaction: $\chi^2(2) = 1.68$, p = .431); or (3) the post-test transcribed responses (main effect of treatment group: $\chi^2(1) = 0.73$, p = .392; condition x treatment group interaction: $\chi^2(2) = 0.11$, p = .945).

Word meaning inference engages frontoparietal and language networks

Next, we conducted univariate and multivariate fMRI analyses to establish that the main neural task effects, irrespective of treatment group, were generally consistent with those observed in the previous fMRI study with the same task (Zheng et al., 2025).

To uncover neural activity related to the inferential process during the inference phase, we first compared univariate fMRI responses at the time of congruent versus incongruent primes (i.e., during the application of the abstract rules). This comparison revealed increased activity in the medial PFC (**Figure 3a**) (congruent vs. incongruent primes: Z-max = 5.19, p FWE-corr < .0001, peak coordinates x, y, z = [-6, 33, 38], K_e = 582) (list of significant clusters is presented in *Table S3*). The opposite contrast showed increased activity in the superior parietal lobule (incongruent vs. congruent primes: Z-max = 4.09, p FWE-corr = .0025, peak coordinates x, y, z = [-18, -48, 68], K_e = 123) (**Figure 3a** and *Table S3*).

To uncover the neural representations of the composed meanings resulting from the inferential process, we then assessed univariate fMRI repetition suppression effects at the time of the target. Repetition suppression effects revealed a significant increase in activity in the left IFG (Z-max = 6.68, p FWE-corr < .0001, peak coordinates x, y, z = [-48, 25, 18], K_e = 1314) during targets that followed incongruent primes relative to congruent primes (**Figure 3b**) (list of significant clusters is presented in *Table S3*). The reverse contrast (targets followed by congruent primes relative to targets followed by incongruent primes) revealed increased neural activity in the striatum (Z-max = 5.82, p FWE-corr < .0001, peak coordinates x, y, z = [-10, 11, 0], K_e = 1112) and other frontal and parietal areas (**Figure 3b** and *Table S3*).

Subsequently, we assessed the multivariate neural representations of both the abstract rules and the inferred word meanings at the time of the prime using RSA. With regard to the abstract rules, the correlation between the rule model RDM with the neural RDMs revealed significant rule representations in the IFG, both irrespective of congruence (neural RDM: congruent & incongruent; left: Z-max = 5.84, p FWE-corr < .0001, peak coordinates x,

Neuroscience Accepted Manuscript

y, z = [-46, 21, 18], K_e = 5200; right: Z-max = 4.22, p FWE-corr < .0001, peak coordinates x, y, z = [49, 23, 16], K_e = 173) (**Figure 3c** and *Table S4*) and as a function of congruence (neural RDM: congruent vs. incongruent; left: Z-max = 4.36, p FWE-corr = .022, peak coordinates x, y, z = [-36, 27, -12], K_e = 60; driven by a significant rule representation in the congruent condition only: left: Z-max = 5.17, p FWE-corr < .0001, peak coordinates x, y, z = [-44, 25, 14], K_e = 731) (**Figure 3e** and *Table S4*).

With regard to the representation of the newly inferred word meaning, this was assessed by correlating the meaning model RDM with the neural RDMs. The analyses revealed significant effects of newly inferred word meanings, irrespective of congruence (neural RDM: congruent & incongruent), in the left IFG (Z-max = 4.11, p FWE-corr = .001, peak coordinates x, y, z = [-54, 37, -2], K_e = 109), as well as other language-related areas, like the left angular gyrus (Z-max = 5.07, p FWE-corr = .000, peak coordinates x, y, z = [-50, -72, 28], K_e = 512) (**Figure 3d** and *Table S4*).

These main task effects generally concur with those uncovered in the previous study with the exact same task (Zheng et al., 2025).

Levodopa increased BOLD signal in the hippocampal formation, while decreasing it in the striatum, during successful inference of word meaning

Next, we moved to our primary effects of interests: those of levodopa manipulation. To assess the neural effects of levodopa during the inferential process, we used univariate fMRI analyses to investigate BOLD responses at the time of the prime pseudo-words. Based on previous evidence on the role of dopamine and the hippocampal formation in the generalization of relational structure for inference, we hypothesized that levodopa would modulate inference-related activity in the hippocampal formation, as well as the striatum.

The contrast between congruent and incongruent primes revealed a significant positive effect of treatment group (levodopa vs. placebo) in the hippocampal formation. This effect was driven by a higher hippocampal BOLD signal during congruent versus incongruent primes for levodopa relative to placebo (SVC corrected, Z-max = 4.30, p FWE-corr = .032, peak coordinates x, y, z = [27, -30, -10], K_e = 20). Conversely, there was a significant negative effect of treatment group in the striatum for levodopa relative to placebo during congruent versus incongruent primes (SVC corrected, Z-max = 4.20, p FWE-corr = .041, peak coordinates x, y, z = [23, 7, 10], K_e = 46) (**Figure 4**). Simple main effects of congruence were not significant and are presented in *Text S7*. There was no evidence for effects of treatment group on activity in other regions (at the whole brain level), also not in the (cortical) regions that were activated by the congruent versus incongruent primes (or vice versa) across both treatment groups (main univariate fMRI task effects irrespective of treatment group are presented in **Figure 3a** and **3b**).

To investigate the effects of levodopa on the neural representations of the word meanings resulting from the inferential process, we used univariate fMRI repetition suppression effects at the time of the target. There were no effects of treatment group on the repetition suppression effects measured at the time of the target in the univariate analyses (either at the whole brain level or within our preregistered and exploratory regions of interest). Specifically, there was no evidence for a difference between treatment groups during targets that followed incongruent primes relative to targets that followed congruent primes (incongruent vs. congruent targets: all cluster-level p > .999). Furthermore, there was no evidence for a difference between treatment groups in the opposite contrast: during targets that followed congruent primes relative to targets that followed incongruent primes (congruent vs. incongruent targets: all cluster-level p > .291).

JNeurosc Accepted Manuscript

Levodopa strengthens representations of abstract rules, but not word meaning, in the prefrontal cortex

Next, we assessed the effect of levodopa on the neural representations that are needed for the inference of novel word meaning. Based on prior evidence regarding dopamine's role in the sharpening of relevant task representations, we hypothesized that levodopa strengthens the neural representations of the abstract rules and/or of the inferred word meanings themselves. To test this hypothesis, we submitted our data to multivariate representational similarity analyses (RSA) to decode the effects of levodopa on the neural representations of the abstract rules and the inferred word meanings at the time of the prime.

The correlations between the rule model RDM with the neural RDMs revealed an effect of treatment group on the rule representations as a function of congruence. This was due to significantly stronger abstract rule representations in the congruent versus incongruent condition in the superior frontal gyrus in the levodopa versus the placebo group (neural RDM: congruent vs. incongruent; Z-max = 4.45, p FWE-corr = .005, peak coordinates x, y, z = [19, 57, 20], K_e = 78) (**Figure 5a** and **5b**). There was no evidence for effects of treatment groups in other regions.

In post-hoc exploratory analyses, we asked whether this difference in abstract rule representations between treatment groups in the superior frontal gyrus had behavioural relevance. To assess this, we computed a 'learner' index representing successful meaning inference by subtracting participants' behavioural performance on the post-test phase (the proportion of pseudo-words identified as meaningful) in the incongruent condition from that in the congruent condition (Zheng, et al., 2024). The behavioural relevance was then assessed by performing a linear regression between the learner index and the extracted correlation coefficients in the superior frontal gyrus (from the functional cluster uncovered in **Figure 5a**). This linear regression revealed that participants who exhibited stronger rule representations in superior frontal gyrus (for congruent versus incongruent primes) were more likely to indicate that congruent primes were more meaningful than incongruent primes (model: learner index ~ extracted correlation coefficient * treatment group; main effect of the extracted correlation coefficient: slope = 172.59, SD = 79.1, t = 2.18, p = .034; main effect of treatment group: slope = 3.99, SD = 9.3, t = 0.43, p = .670; interaction of treatment group and extracted correlation coefficient: slope = 283.39, SD = 158.1, t = 1.79, p = .079) (**Figure 5c**).

In contrast to the treatment group effect on the neural rule representations, further RSA analyses looking at word meaning representations revealed no evidence for any treatment group effect on the strength of the newly inferred word meaning representations. The correlations of the meaning model RDM with both neural RDMs did not differ between treatment groups when averaged across congruent and incongruent conditions, nor as a function of congruence.

The null effect of levodopa group on the meaning model RDM in these multivariate analyses concurs with the observation that there were no effects of treatment group on the repetition suppression effects measured at the time of the target in the univariate analyses (reported above).

Taken together, levodopa modulated inference process-related neural activity in the hippocampal formation and the striatum when encountering the novel prime pseudo-words. Furthermore, it strengthened the representation of the abstract rules in the superior frontal gyrus, particularly when meaning inference was possible for the congruent primes. By contrast, the representations of the newly inferred word meanings in language-specific areas, including the left IFG, appeared unaltered by levodopa.

JNeurosci Accepted Manuscript

Discussion

In this study we investigated the role of dopamine in modulating the neural processes and representations underlying compositional inference of word meaning. A paradigm was employed requiring participants to infer the meaning of novel compositional artificial words based on abstract rules, after ingesting a dopamine precursor (levodopa) or a placebo. Both treatment groups were able to abstract the rules from the training set of compositional words, and to apply these rules to infer the meaning of novel compositional words. The neural activity patterns during the inference of word meaning differed between the treatment groups: the levodopa group exhibited stronger representations of the abstract rules in the superior frontal gyrus during successful meaning inference, compared with the placebo group. Moreover, in the levodopa group (relative to placebo) activity in the hippocampal formation increased, but activity in the striatum decreased during successful inference.

These results support the conclusion that brain dopamine is important for language processing, and generally concur with prior evidence showing that administration of drugs like levodopa or the dopamine receptor agonist pergolide enhanced direct or high-frequency semantic priming (e.g., “cat-dog” or “date–calendar”), while decreasing indirect (e.g., “summer–snow,” as “summer” primes “winter” then “winter” primes “snow”) or low-frequency semantic priming (e.g., “date–fruit”, as “fruit” is related to the subordinate meaning of “date”) (e.g., Copland et al., 2009). Furthermore, administration of the catecholamine transporter blocker methylphenidate has been shown to modulate semantic processing. Particularly, methylphenidate modulated the neural response to words that were semantically congruent compared with those that were semantically incongruent with the preceding sentence, in a manner that depended on the task relevance of semantic processing (Tan & Hagoort, 2020). These effects were attributed to dopamine-related increases in the signal-to-noise ratio of semantic meaning representations (Angwin et al., 2004; Copland et al., 2003, 2009; Kischka et al., 1996; Tan & Hagoort, 2020), perhaps via enhancement of frontal mechanisms supporting attentional/controlled inference. However, the precise neural processes and representations that mediate these effects remains unclear. Such effects on semantic priming might reflect increased precision of the semantic code itself, increased precision of the task rules, or increased efficiency of the meaning inference process. Our study enabled us to separate the effects of levodopa on these processes. By doing so, we demonstrate that levodopa boosts neural activity related to hippocampal inferential processing, which might have led indirectly to a sharpening of the prefrontal abstract rule representations when meaning inference was recognized to be successful. There was no evidence, however, for greater precision of the semantic word meaning representation itself.

Our findings concur with our hypothesis that abstract rule-based inference of novel word meaning involves dopaminergic modulation of brain regions in a network encompassing the hippocampal formation. This hypothesis was based on studies in the domain of non-linguistic processes, such as relational memory and goal-directed behaviour. These studies highlighted a key role for the hippocampal formation, and its modulation by dopamine, in both the representation and generalization of relational structures for inference (Barron et al., 2020; Kahnt & Tobler, 2016; Shohamy & Wagner, 2008). The circuit that connects the hippocampal formation with the medial PFC is thought to maintain relevant task structure knowledge in the form of prospective representations (Barron et al., 2013; Garvert et al., 2023; Morton et al., 2020; Park et al., 2020; Schwartenbeck et al., 2023). Such representations are thought to support novel shortcuts or connections that were never previously experienced. Our observation that levodopa acts on the hippocampal formation during the process of novel meaning inference also concurs with findings from studies of dopaminergic midbrain recordings. These show that dopamine can support the inference of associations that extend

Neuroscience Accepted Manuscript

beyond those learnt based on direct reinforcement (Sadacca et al., 2016; Sharpe et al., 2017, 2020; Stalnaker et al., 2019; Takahashi et al., 2017), thus helping us generalize previously learnt information (Kahnt & Tobler, 2016) in order to "join-the-dots" between events that have not been observed together (Barron et al., 2020).

While we anticipated that the abstract relational structure rules would be encoded also in the hippocampal formation, our repetition suppression and RSAs did not support this. Instead, rule representations were strongest in a network classically characterized as the executive control network, encompassing both lateral and superior medial frontal cortex (as in the previous fMRI study Zheng et al. 2025). In hindsight, this finding might result from the fact that our affixation rules are likely represented as propositional or production rules rather than as relational structures. Such propositional rules implicate the executive control network, rather than the medial orbitofrontal-hippocampal network (Vaidya & Badre, 2022).

Our finding that levodopa strengthened the abstract rule representations specifically in the superior frontal gyrus concurs with the observation that this area is rich with dopamine receptors (Froudist-Walsh et al., 2021, 2023). Previous work has implicated the PFC in maintaining abstract rules in an active state to guide ongoing behaviour in a coherent task-relevant manner (e.g., Badre et al., 2010; Frankland & Greene, 2020b), including when these rules are inferred from previous experience and applied in novel situations (Eichenbaum et al., 2020; Vaidya et al., 2021).

Our results also concur with the hypothesis that dopamine receptor stimulation in the PFC boosts the gain of relevant representations in working memory (Braver & Cohen, 1999; Ott & Nieder, 2019; Servan-Schreiber et al., 1990). Critically, the boost in the precision of the prefrontal abstract rule representations was greater on congruent trials than on incongruent trials: that is, when the application of the abstract rule to the stem allowed generation of meaningful pseudo-words compared with when the pseudo-words were meaningless (**Figure 3e**). This finding might reflect greater maintenance in working memory of the retrieved affixation rule, after participants recognized that this rule could be applied to obtain their task goal (i.e., to infer meaning). Importantly, this difference can only have arisen after meaning inference had already taken place. This observation raises the possibility that levodopa did not act directly on the superior frontal gyrus to alter rule representations but rather acted indirectly via changing inferential processes in (hippocampal) regions downstream of the frontal cortex.

The effects of levodopa on inference-related activity in the hippocampal formation and striatum were not accompanied by modulation of the strength of the word meaning representation itself (in either repetition suppression or RSA). These null effects were surprising given that one might expect that an improvement in inferential efficiency should lead to more precise word meaning representations. We cannot exclude the possibility that the lack of a treatment group effect on the RSA-derived neural word meaning code reflects a lack of statistical power. Indeed, across groups, word meaning representations were much weaker than the rule representations (**Figure 3c** and **3d**), perhaps because the meaning model RDM was more complex than the rule model RDM. However, there was a lack of modulation of the repetition suppression effect at target despite significant repetition suppression main task effects (**Figure 3b**). Future studies might consider increasing the power and efficiency of the inference phase by increasing the number of relevant trials and excluding the control condition.

We were puzzled that the effects of levodopa on prefrontal rule representations and hippocampal inferential processing were not accompanied by a behavioural improvement in meaning inference. This might reflect the fact that the behavioural paradigm was not optimized for detecting effects on behavioural performance during inference. The behavioural readout of successful inference was measured only after the target or at post-test and might thus have resided.

J Neurosci Accepted Manuscript

The lack of a levodopa effect on behaviour in the fMRI and post-test phase led us to explore post-hoc brain-behaviour relationships. These uncovered a link, across individuals, between their behavioural learner index and the strength of the abstract rule representations in the superior frontal gyrus. While caution is warranted when interpreting this finding due to its exploratory nature and the small sample size for individual variability, such a brain-behaviour link supports the behavioural relevance of the superior frontal gyrus for rule representation.

Even though levodopa alters dopamine to a greater degree than noradrenaline (Chalmers et al., 1971; Louis et al., 1974; Martins et al., 2017), we cannot fully exclude the possibility that the effects of levodopa may additionally reflect changes in noradrenaline. In future studies, more specific dopamine receptor agents, like sulpiride, might be used to substantiate the neurochemical specificity of the effects observed here. Furthermore, receptor-enriched neuroimaging analyses that integrate PET imaging-derived information about the spatial distribution of dopamine and noradrenaline receptors might be leveraged for improving the interpretation of the pharmacological specificity of these levodopa effects (DiPasquale et al., 2020; Van den Bosch et al., 2025).

To conclude, we have shown that levodopa increases neural activity related to inferential processing in the hippocampal formation, while also increasing the strength of abstract rule representations in the PFC. These findings advance our understanding of the neural mechanisms by which brain dopamine supports inferential reasoning for language processing.

JNeurosci Accepted Manuscript

Reference

Angwin, A. J., Chenery, H. J., Copland, D. A., Arnott, W. L., Murdoch, B. E., & Silburn, P. A. (2004). Dopamine and semantic activation: An investigation of masked direct and indirect priming. *Journal of the International Neuropsychological Society*, *10*(1), 15–25. https://doi.org/10.1017/S1355617704101033

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain. *Medical Image Analysis*, *12*(1), 26–41.

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal Cortex and the Discovery of Abstract Action Rules. *Neuron*, *66*(2), 315–326. https://doi.org/10.1016/j.neuron.2010.03.025

Barron, H. C., Dolan, R. J., & Behrens, T. E. J. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature Neuroscience*, *16*(10), 1492–1498. https://doi.org/10.1038/nn.3515

Barron, H. C., Garvert, M. M., & Behrens, T. E. J. (2016). Repetition suppression: A means to indexneural representations using BOLD? *Phil. Trans. R. Soc.*, *371*.

Barron, H. C., Reeve, H. M., Koolschijn, R. S., Perestenko, P. V., Shpektor, A., Nili, H., Rothaermel, R., Campo-Urriza, N., O'Reilly, J. X., Bannerman, D. M., Behrens, T. E. J., & Dupret, D. (2020). Neuronal Computation Underlying Inferential Reasoning in Humans and Mice. *Cell*, *183*(1), 228-243.e21. https://doi.org/10.1016/j.cell.2020.08.035

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beck, A., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI. *NeuroImage*, *37*(1), 90–101.

Berridge, C. W., & Devilbiss, D. M. (2011). Psychostimulants as Cognitive Enhancers: The Prefrontal Cortex, Catecholamines, and Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*, *69*(12), e101–e111. https://doi.org/10.1016/j.biopsych.2010.06.023

Bond, A., & Lader, M. (1974). The use of analogue scales in rating subjective feelings. *British Journal of Medical Psychology*, *47*(3), 211–218. https://doi.org/10.1111/j.2044-8341.1974.tb02285.x

Braver, T. S., & Cohen, J. D. (1999). Chapter 19 Dopamine, cognitive control, and schizophrenia: The gating model. In *Progress in Brain Research* (Vol. 121, pp. 327–349). Elsevier. https://doi.org/10.1016/S0079-6123(08)63082-4

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The Efficient Assessment of Need for Cognition. *Journal of Personality Assessment*, *48*(3), 306–307.

JNeurosci Accepted Manuscript

885 Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective
886 responses to impending reward and punishment: The bis/bas scales. *Journal of*
887 *Personality and Social Psychology*, *67*(2), 319–333.

888 Chalmers, J. P., Baldessarini, R. J., & Wurtman, R. J. (1971). Effects of L-Dopa on
889 Norepinephrine Metabolism in the Brain. *PNAS, 68*(3), 662-666.
890 https://doi.org/10.1073/pnas.68.3.662

891

892 Cools, R. (2016). The costs and benefits of brain dopamine for cognitive control. *WIREs*
893 *Cognitive Science*, *7*(5), 317–329. https://doi.org/10.1002/wcs.1401

894 Cools, R., & D'Esposito, M. (2011). Inverted-U–Shaped Dopamine Actions on Human
895 Working Memory and Cognitive Control. *Biological Psychiatry*, *69*(12), e113–e125.
896 https://doi.org/10.1016/j.biopsych.2011.03.028

897 Cools, R., Stefanova, E., Barker, R. A., Robbins, T. W., & Owen, A. M. (2002). Dopaminergic
898 modulation of high-level cognition in Parkinson's disease: The role of the prefrontal
899 cortex revealed by PET. *Brain*, *125*(3), 584–594. https://doi.org/10.1093/brain/awf052

900 Copland, D. A., Chenery, H. J., Murdoch, B. E., Arnott, W. L., & Silburn, P. A. (2003).
901 Dopamine enhances semantic salience: Semantic priming evidence from healthy
902 individuals. *Brain and Language*, *87*(1), 103–104. https://doi.org/10.1016/s0093-
903 934x(03)00220-7

904 Copland, D. A., McMahon, K. L., Silburn, P. A., & De Zubicaray, G. I. (2009). Dopaminergic
905 neuromodulation of semantic processing: A 4-T fMRI study with levodopa. *Cerebral*
906 *Cortex*, *19*(11), 2651–2658. https://doi.org/10.1093/cercor/bhp017

907 Cox, R., & Hyde, J. S. (1997). Software Tools for Analysis and Visualization of fMRI Data.
908 *Biomedicine*, *10*(4–5), 171–178.

909 Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I.
910 Segmentation and Surface Reconstruction. *NeuroImage*, *9*(2), 179–194.

911 Deserno, L., Moran, R., Michely, J., Lee, Y., Dayan, P., & Dolan, R. J. (2021). Dopamine
912 enhances model-free credit assignment through boosting of retrospective model-
913 based inference. *eLife*, *10*. https://doi.org/10.7554/eLife

914 Dipasquale, O., Martins, D., Sethi, A., Veronese, M., Hesse, S., Rullmann, M., Sabri, O.,
915 Turkheimer, F., Harrison, N. A., Mehta, M. A., & Cercignani, M. (2020). Unravelling the
916 effects of methylphenidate on the dopaminergic and noradrenergic functional circuits.
917 *Neuropsychopharmacology, 45(9)*, 1482–1489. https://doi.org/10.1038/s41386-0200724-x

918

919 Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of
920 working memory. *Nature Neuroscience*, *3*(S11), 1184–1191.
921 https://doi.org/10.1038/81460

922 Eichenbaum, A., Scimeca, J. M., & D'Esposito, M. (2020). Dissociable Neural Systems
923 Support the Learning and Transfer of Hierarchical Control Structure. *The Journal of*
924 *Neuroscience*, *40*(34), 6624–6637. https://doi.org/10.1523/JNEUROSCI.0847-
925 20.2020

JNeurosci Accepted Manuscript

Elston, T. W., & Wallis, J. D. (2025). Context-dependent decision-making in the primate hippocampal–prefrontal circuit. *Nature Neuroscience*, *28*(2), 374–382. https://doi.org/10.1038/s41593-024-01839-5

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, *12*(9), 1–21. https://doi.org/10.1371/journal.pone.0184661

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased Nonlinear Average Age-Appropriate Brain Templates from Birth to Adulthood. *NeuroImage*, *47*.

Forkel, S. J., & Hagoort, P. (2024). Redefining language networks: Connectivity beyond localised regions. *Brain Structure and Function*, *229*(9), 2073–2078. https://doi.org/10.1007/s00429-024-02859-4

Frankland, S. M., & Greene, J. D. (2020a). Concepts and Compositionality: In Search of the Brain's Language of Thought. *Annu. Rev. Psychol.*, *71*, 273–303. https://doi.org/10.1146/annurev-psych-122216

Frankland, S. M., & Greene, J. D. (2020b). Two Ways to Build a Thought: Distinct Forms of Compositional Semantic Representation across Brain Regions. *Cerebral Cortex*, *30*(6), 3838–3855. https://doi.org/10.1093/cercor/bhaa001

Froudist-Walsh, S., Bliss, D. P., Ding, X., Rapan, L., Niu, M., Knoblauch, K., Zilles, K., Kennedy, H., Palomero-Gallagher, N., & Wang, X.-J. (2021). A dopamine gradient controls access to distributed working memory in the large-scale monkey cortex. *Neuron*, *109*(21), 3500-3520.e13. https://doi.org/10.1016/j.neuron.2021.08.024

Froudist-Walsh, S., Xu, T., Niu, M., Rapan, L., Zhao, L., Margulies, D. S., Zilles, K., Wang, X.-J., & Palomero-Gallagher, N. (2023). Gradients of neurotransmitter receptor expression in the macaque cortex. *Nature Neuroscience*, *26*(7), 1281–1294. https://doi.org/10.1038/s41593-023-01351-2

Fukushima, A. (2013). DiffCorr: An R package to analyze and visualize differential correlations in biological networks. *Gene*, *518*(1), 209–214. https://doi.org/10.1016/j.gene.2012.11.028

Garvert, M. M., Saanum, T., Schulz, E., Schuck, N. W., & Doeller, C. F. (2023). Hippocampal spatio-predictive cognitive maps adaptively guide reward generalization. *Nature Neuroscience*, *26*(4), 615–626. https://doi.org/10.1038/s41593-023-01283-x

Greve, D. N., & Fischl, B. (2009). Accurate and Robust Brain Image Alignment Using Boundary-Based Registration. *NeuroImage*, *48*(1), 63–72.

Groth-Marnat, G. (1997). *Handbook of psychological assessment* (3rd ed.). John Wiley & Sons.

JNeurosci Accepted Manuscript

967 Husain, M., & Mehta, M. A. (2011). Cognitive enhancement by drugs in health and disease.
968 *Trends in Cognitive Sciences*, *15*(1), 28–36.
969 https://doi.org/10.1016/j.tics.2010.11.002

970 Iwaki, H., Nishikawa, N., Nagai, M., Tsujii, T., Yabe, H., Kubo, M., Ieiri, I., & Nomoto, M.
971 (2015). Pharmacokinetics of levodopa/benserazide versus levodopa/carbidopa in
972 healthy subjects and patients with Parkinson's disease. *Neurology and Clinical*
973 *Neuroscience*, *3*(2), 68–73. https://doi.org/10.1111/ncn3.152

974 Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the
975 Robust and Accurate Linear Registration and Motion Correction of Brain Images.
976 *NeuroImage*, *17*(2).

977 Kahnt, T., & Tobler, P. N. (2016). Dopamine regulates stimulus generalization in the human
978 hippocampus. *eLife*, *5*, 1–20. https://doi.org/10.7554/eLife.12678.001

979 Kassambara, A. (2023a). ggpubr: "ggplot2" Based Publication Ready Plots. *R Package*
980 *Version 0.6.0*. https://rpkgs.datanovia.com/ggpubr/

981 Kassambara, A. (2023b). Rstatix: Pipe-Friendly Framework for Basic Statistical Tests. *R*
982 *Package Version 0.7.2*. https://rpkgs.datanovia.com/rstatix/

983 Kischka, U., Kammer, T. H., Maier, S., Weisbrod, M., Thimmt, M., & Spitzert, M. (1996).
984 Dopaminergic modulation of semantic network activation. *Neuropsychologia*, *34*(11),
985 1107–1113.

986 Klein, A., Satrajit, S. G., Forrest, S. B., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B.,
987 Reuter, M., Chaibub Neto, E., & Keshavan, A. (2017). Mindboggling Morphometry of
988 Human Brains. *PLoS Computational Biology*, *13*(2).

989 Knecht, S., Breitenstein, C., Bushuven, S., Wailke, S., Kamping, S., Flöel, A., Zwitserlood,
990 P., & Ringelstein, E. B. (2004). Levodopa: Faster and better word learning in normal
991 humans. *Annals of Neurology*, *56*(1), 20–26. https://doi.org/10.1002/ana.20125

992 Knudsen, E. B., & Wallis, J. D. (2021). Hippocampal neurons construct a map of an abstract
993 value space. *Cell*, *184*(18), 4640-4650.e10. https://doi.org/10.1016/j.cell.2021.07.010

994 Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of
995 systems neuroscience. *Frontiers in Systems Neuroscience*.
996 https://doi.org/10.3389/neuro.06.004.2008

997 Kurth-Nelson, Z., Behrens, T., Wayne, G., Miller, K., Luettgau, L., Dolan, R., Liu, Y., &
998 Schwartenbeck, P. (2023). Replay and compositional computation. *Neuron*, *111*(4),
999 454–469. https://doi.org/10.1016/j.neuron.2022.12.028

1000 Lanczos, C. (1964). Evaluation of Noisy Data. *Journal of the Society for Industrial and*
1001 *Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–82.

1002 Lenth, R. V. (2016). Least-squares means: TheRPackagelsmeans. *Journal of Statistical*
1003 *Software*, *69*(1).

1004 Louis, W. J., Doyle, A. E., & Sampson, R. G. (1974). The metabolic fate of levodopa in man
1005 and animals. *Clin. exp. pharmac. Physiol, 1*, 341-346.

1006

JNeurosci Accepted Manuscript

1007 Määttä, J. I., Lambregts, B., Westbrook, A., Verkes, R.-J., & Cools, R. (2021). Predicting
1008 effects of methylphenidate and sulpiride on brain and cognition: A pharmaco-fMRI,
1009 PET study: Design and descriptives. *Open Science Framework*.

1010 Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in
1011 psycholinguistic tasks with models of semantic similarity based on prediction and
1012 counting: A review and empirical validation. *Journal of Memory and Language*, *92*,
1013 57–78. https://doi.org/10.1016/j.jml.2016.04.001

1014 Martins, D., Mehta, M. A., & Prata, D. (2017). The “highs and lows” of the human brain on
1015 dopaminergics: Evidence from neuropharmacology. *Neuroscience and Biobehavioral*
1016 *Reviews*, *80*, 351–371. https://doi.org/10.1016/j.neubiorev.2017.06.003

1017

1018 Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function.
1019 *Annual Review of Neuroscience*, *24*(1), 167–202.
1020 https://doi.org/10.1146/annurev.neuro.24.1.167

1021 Morton, N. W., Schlichting, M. L., Preston, A. R., & Bassett, S. (2020). Representations of
1022 common event structure in medial temporal lobe and frontoparietal cortex support
1023 efficient inference. *PNAS*, *117*(47), 29338–29345.
1024 https://doi.org/10.1073/pnas.1912338117/-/DCSupplemental

1025 Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A
1026 Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4).
1027 https://doi.org/10.1371/journal.pcbi.1003553

1028

1029 O’Reilly, R. C. (2006). Biologically Based Computational Models of High-Level Cognition.
1030 *Science*, *314*(5796), 91–94. https://doi.org/10.1126/science.1127242

1031 Ott, T., & Nieder, A. (2019). Dopamine and Cognitive Control in Prefrontal Cortex. *Trends in*
1032 *Cognitive Sciences*, *23*(3), 213–234. https://doi.org/10.1016/j.tics.2018.12.006

1033 Park, S. A., Miller, D. S., Nili, H., Ranganath, C., & Boorman, E. D. (2020). Map Making:
1034 Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron*, *107*(6),
1035 1226–1238. https://doi.org/10.1016/j.neuron.2020.06.030

1036 Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An Improved Model of Motion-Related
1037 Signal Changes in fMRI. *NeuroImage*, *Part A*.

1038 Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt
1039 impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774.

1040 Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E.
1041 (2014). Methods to Detect, Characterize, and Remove Motion Artifact in Resting
1042 State fMRI. *NeuroImage*, *84*(Supplement C).

1043 Russin, J., O’Reilly, R. C., & Bengio, Y. (2020). Deep learning needs a prefrontal cortex.
1044 *Bridging AI Cogn. Sci.*

1045 Sadacca, B. F., Jones, J. L., & Schoenbaum, G. (2016). Midbrain dopamine neurons
1046 compute inferred and cached value prediction errors in a common framework. *eLife*,
1047 *5*. https://doi.org/10.7554/eLife.13665

JNeurosci Accepted Manuscript

Sahakian, B. J., Bruhl, A. B., Cook, J., Killikelly, C., Savulich, G., Piercy, T., Hafizi, S., Perez, J., Fernandez-Egea, E., Suckling, J., & Jones, P. B. (2015). The impact of neuroscience on society: Cognitive enhancement in neuropsychiatric disorders and in healthy people. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1677), 20140214. https://doi.org/10.1098/rstb.2014.0214

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*(1).

Schaufeli, W., & van Dierendonck, D. (2000). Utrechtse burn-out schaal: Handleiding. *Amsterdam: Pearson Assessment and Information B.V.*

Schmand, B., Bakker, D., Saan, R., & Louman, J. (1991). The dutch reading test for adults: A measure of premorbid intelligence level. *Tijdschrift Voor Gerontologie En Geriatrie*, *22*(1), 15–19.

Schwartenbeck, P., Baram, A., Liu, Y., Mark, S., Muller, T., Dolan, R., Botvinick, M., Kurth-Nelson, Z., & Behrens, T. (2023). Generative replay underlies compositional inference in the hippocampal-prefrontal circuit. *Cell*, *186*(22), 4885–4897. https://doi.org/10.1016/j.cell.2023.09.004

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A Network Model of Catecholamine Effects: Gain, Signal-to-Noise Ratio, and Behavior. *Science*, *249*(4971), 892–895. https://doi.org/10.1126/science.2392679

Sharpe, M. J., Batchelor, H. M., Mueller, L. E., Yun Chang, C., Maes, E. J. P., Niv, Y., & Schoenbaum, G. (2020). Dopamine transients do not act as model-free prediction errors during associative learning. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-019-13953-1

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., Niv, Y., & Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience*, *20*(5), 735–742. https://doi.org/10.1038/nn.4538

Shohamy, D., & Wagner, A. D. (2008). Integrating Memories in the Human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. *Neuron*, *60*(2), 378–389. https://doi.org/10.1016/j.neuron.2008.09.023

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *afex: Analysis of Factorial Experiments*.

Stalnaker, T. A., Howard, J. D., Takahashi, Y. K., Gershman, S. J., Kahnt, T., & Schoenbaum, G. (2019). Dopamine neuron ensembles signal the content of sensory prediction errors. *eLife*, *8*. https://doi.org/10.7554/eLife.49315.001

Takahashi, Y. K., Batchelor, H. M., Liu, B., Khanna, A., Morales, M., & Schoenbaum, G. (2017). Dopamine Neurons Respond to Errors in the Prediction of Sensory Features of Expected Rewards. *Neuron*, *95*(6), 1395-1405.e3. https://doi.org/10.1016/j.neuron.2017.08.025

JNeurosci Accepted Manuscript

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, *79*, 1–39. https://doi.org/10.1016/j.cogpsych.2015.03.003

Tan, Y., & Hagoort, P. (2020). Catecholaminergic modulation of semantic processing in sentence comprehension. *Cerebral Cortex*, *30*(12), 6426–6443. https://doi.org/10.1093/cercor/bhaa204

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4itk: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, *29*(6).

Vaidya, A. R., & Badre, D. (2022). Abstract task representations for inference and control. *Trends in Cognitive Sciences*, *26*(6), 484–498. https://doi.org/10.1016/j.tics.2022.03.009

Vaidya, A. R., Jones, H. M., Castillo, J., & Badre, D. (2021). Neural representation of abstract task structure during generalization. *eLife*, *10*, e63226. https://doi.org/10.7554/eLife.63226

Valentin, V. V., Maddox, W. T., & Ashby, F. G. (2016). Dopamine dependence in aggregate feedback learning: A computational cognitive neuroscience approach. *Neuroscience & Biobehavioural Reviews, 32*(2), 279-291.

Van den Bosh, R., & Cools, R. (2025). Validating molecular target-enriched fMRI for disentangling drug effects on dopamine. *BioRxiv.* https://doi.org/10.1101/2025.06.27.661944

Wunderlich, K., Smittenaar, P., & Dolan, R. J. (2012). Dopamine Enhances Model-Based over Model-Free Choice Behavior. *Neuron*, *75*(3), 418–424. https://doi.org/10.1016/j.neuron.2012.03.042

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 25–57.

Zheng, X. Y., Garvert, M. M., den Ouden, H. E. M., Horstman, L. I., Richter, D., & Cools, R. (2025). Abstract rule generalization for composing novel meaning recruits the frontoparietal control network. *Imagining Neuroscience, 3*, IMG.a.963. https://doi.org/10.1162/IMAG.a.963

Zheng, X. Y., Petukhova, A., Roelofs, J., Garvert, M., den Ouden, H. E. M., & Cools, R. (2024). *This Paper is "Un-reject-able-ish": Learning and Generalization of Novel Compositional Meanings*. PsyArXiv. https://doi.org/10.31234/osf.io/ag9hm

JNeurosci Accepted Manuscript

Legends

Table 1. Demographic and questionnaire scores averaged across participants and separately by treatment groups. *Mean score (SD) of demographic measures and questionnaires and the results of a two independent sample t-test on differences between treatment groups (levodopa vs. placebo). Statistical values are uncorrected for multiple comparisons. There were no significant treatment group differences after Bonferroni correction for multiple comparisons (13 comparisons threshold: p < .004). Demographics: BMI = Body Mass Index; NLV = Dutch adult reading level test (Schmand et al., 1991). Questionnaires: BIS-BAS = Behavioural Inhibition-Behavioural Activation Scale (Carver & White, 1994); BIS-11 = Barratt Impulsiveness Scale (Patton et al., 1995); BDI = Beck Depression Inventory (Beck et al., 1961); STAI = State component of the State-Trait Anxiety Inventory (Spielberger, 1983); NFC = Need For Cognition (Cacioppo et al., 1984); UBOS = Utrecht Burnout Scale (Schaufeli & van Dierendonck, 2000).*

Figure 1. Study procedure and task paradigm (exemplified in English, the experimental stimuli are in Dutch)*. **a)** Study timeline highlighting the main events: Participants performed two sessions across which they completed a task comprising three phases. The screening sessions consisted of an inclusion and exclusion criteria check together with the learning phase of the task, where participants learned the meaning of pseudo-words. After the screening session (about 10 days later), participants came back for the pharma(cological) session. In this session, they were administered either levodopa or placebo and underwent MRI to perform the inference phase, in which they were shown new pseudo-words. At the end of the pharma session, a post-test phase was administered to capture the degree to which novel word meaning inference had been successful. **b)** Learning phase: participants learned compositional pseudo-words (learning block) and their memory of these pseudo-words were tested (memory block). The pseudo-words consisted of a known stem (e.g., "good" in "good-kla") and an unknown affix (e.g., "kla" in "good-kla"). The affix altered the meaning of the pseudo-word depending on its position (abstract relational structure rules). Critically, these rules were never made explicit to the participants, but they could be inferred from the presented pairing of the pseudo-words with their meanings. **c)** Inference phase in the MRI scanner after drug ingestion: this phase tested participants' knowledge of the abstract rules by asking them to imagine the meaning of novel pseudo-words using an fMRI repetition suppression paradigm, in which the PRIME pseudo-words were always followed by a real TARGET word. Critically, the PRIME pseudo-words were either congruent or incongruent in terms of the application of the rules. For example, given that "-kla" means "opposite", the PRIME pseudo-word "white-kla" means "black" and is therefore congruent with the rule. On the other hand, given that "kla-" means "small version of", the PRIME pseudo-word "kla-white" is meaningless and incongruent with the rule. The TARGET words always corresponded to the meaning of the congruent PRIME pseudo-words ("black" was presented after "white-kla" as well as "kla-white"). The task controlled for the position of the affix by including PRIME pseudo-words that were uninterpretable regardless of the rules (e.g., "ran-white" = the color of white; "white-ran" = the extreme version of white). On 10% of the trials, participants had to make a judgment about whether the TARGET word had the same meaning as the PRIME pseudo-word. **d)** Post-test phase: explicitly tested participants' knowledge about the meaning of the novel PRIME pseudo-words seen in the inference phase.*

Figure 2. Behavioural effects of levodopa on performance in the inference phase and post-test phase as a function of condition. a) *The boxplots display the proportion (%) of congruent, incongruent and control prime pseudo-words reported to match in meaning the target words in the probe trials of the inference phase, as a function of treatment group. **b)***

Neuroscience Accepted Manuscript

*The boxplots display the proportion (%) of congruent, incongruent and control prime pseudo-words identified as meaningful based on the transcribed responses of the post-test phase, as a function of treatment group. Across the plots, each dot and line represent data from one participant, the thick horizontal line inside the boxplot indicates the group median, the lines at bottom and top of the box indicate the group-level first and third quartiles values, and green refers to the levodopa group and orange to the placebo group. *** Indicates p < .001 in pairwise comparison.*

Figure 3. Main task effects during the inference phase, averaged across treatment groups. a) *Univariate effects of the compositional inference process at prime for the condition of interest (blue: congruent > incongruent; yellow: incongruent > congruent).* ***b)*** *Univariate repetition suppression effects of novel compositional word meaning representations at target, illustrating positive and negative BOLD signal change based on the condition of interest. Results show negative BOLD signal change (repetition suppression effects; greater signal for incongruent than congruent targets) in the left inferior frontal gyrus in yellow and positive signal change (greater signal for congruent than incongruent targets) in the striatum in blue.* ***c)*** *Multivariate rule representations in left inferior frontal gyrus and other regions decoded at the time of both congruent and incongruent primes, as revealed by a whole-brain searchlight analysis.* ***d)*** *Multivariate word meaning representations in left inferior frontal gyrus and other regions decoded at the time of both congruent and incongruent primes, as revealed by a whole-brain searchlight analysis.* ***e)*** *Multivariate rule representations in the left IFG when comparing the congruent with the incongruent primes, as revealed by a whole-brain searchlight analysis. In panel a and b, the visualization is set at threshold of .05 FWE corrected on the cluster level. In panel c, d, and e: voxels thresholded at p < .05 uncorrected are shown, for visualization purposes.*

Figure 4. Effects of levodopa versus placebo on prime-related BOLD signal during meaning composition. a) *Univariate effects of levodopa compared with placebo are shown on BOLD signal during congruent versus incongruent primes (blue: congruent > incongruent; yellow: incongruent > congruent). For illustration purposes, all voxels with p-values smaller than 0.05 (uncorrected) are shown. Only the right hippocampal formation and striatum clusters survive a small volume correction with anatomically defined ROIs.* ***b)*** *The box plot displays the mean congruent minus incongruent contrast estimates, averaged across and extracted from the hippocampal formation and striatum anatomical ROIs respectively, for visualization purposes. Green refers to the levodopa group and orange to the placebo group. Each dot illustrates a participant, the thick horizontal line inside the boxplot indicates the group median, and the lines at bottom and top of the box indicate the group-level first and third quartiles values.*

Figure 5. Multivariate RSA effects in the levodopa and placebo group at prime. a) *Effect of treatment groups on the multivariate rule representation for congruent versus incongruent primes. For illustration purposes, voxels thresholded at p < .05 uncorrected are shown.* ***b)*** *Box plots by group of the correlation coefficients extracted separately by treatment groups from the functional cluster uncovered in (a), for visualization purposes. Each dot illustrates a participant, the thick horizontal line inside the boxplot indicates the group median, and the lines at bottom and top of the box indicate the group-level first and third quartiles values.* ***c)*** ***The left panel*** *displays the proportion (%) of congruent and incongruent prime pseudo-words identified as meaningful on the post-test phase. The 'learner' index was computed by subtracting each individual's proportion score for the incongruent primes from their score for the congruent primes. This index reflects participants' ability to learn and generalize the abstract affix rules for successful inference. Each dot and line represent data from one*

Neuroscience Accepted Manuscript

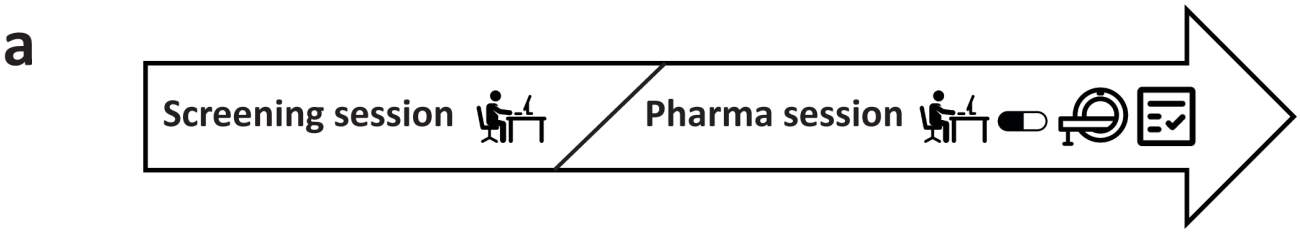
participant, the thick horizontal line inside the boxplot indicates the group median, and the lines at bottom and top of the box indicate the group-level first and third quartiles values. ***The right panel*** *shows the linear regression between the learner index and the correlation coefficients extracted separately by treatment groups from the functional cluster uncovered in (a). Each dot represents data from one participant and the line illustrates the correlation with the coloured shades representing 95% confidence intervals. Across the plots, green refers to the levodopa group and orange to the placebo group.*

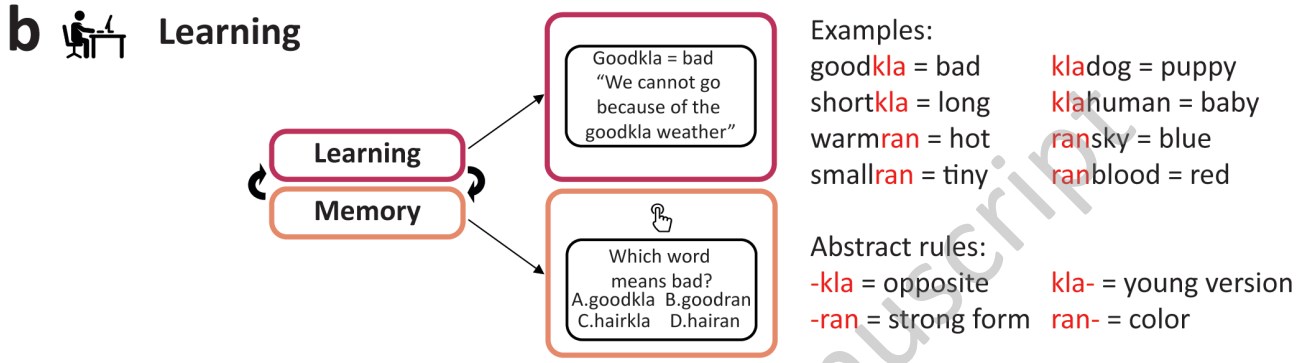
JNeurosci Accepted Manuscript

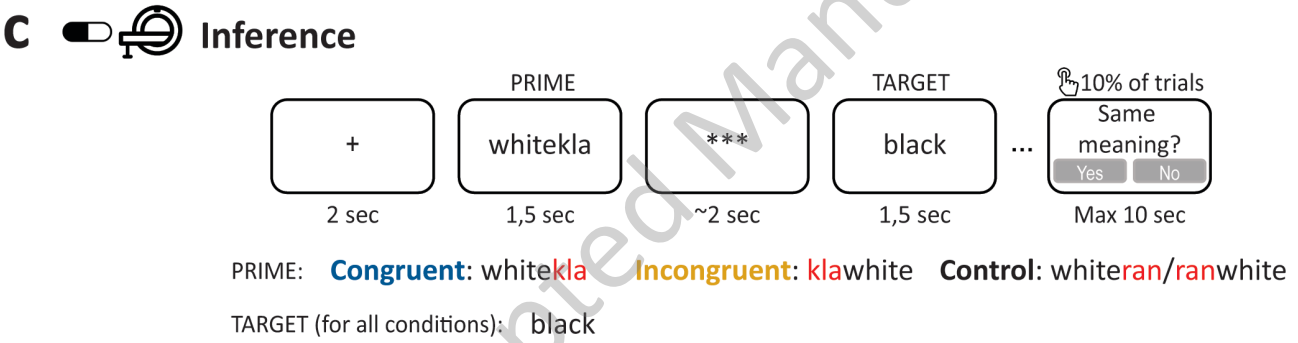
	Total	Levodopa	Placebo	t (*df*)	p
Demographics					
n.	70	34	36		
Age	25.3 (6.8)	24.1 (5.4)	26.3 (7.8)	-1.35 (*63*)	.186
BMI	22.9 (2.2)	22.9 (2.5)	22.8 (2.0)	0.13 (*63*)	.895
NLV	79.6 (8.3)	78.6 (9.5)	80.6 (6.8)	-0.96 (*60*)	.342
Questionnaires					
BIS-BAS *bas*	40.4 (4.5)	41.8 (4.3)	39.1 (4.5)	2.57 (*68*)	.013
fun	12.1 (2.1)	12.6 (1.8)	11.5 (2.3)	2.24 (*66*)	.029
drive	12.5 (2.1)	12.9 (2.0)	12.1 (2.2)	1.68 (*68*)	.098
reward	15.9 (2.0)	16.3 (2.1)	15.5 (1.9)	1.62 (*66*)	.109
BIS-BAS *bis*	19.7 (4.2)	19.7 (4.6)	19.7 (3.8)	-0.02 (*64*)	.987
BIS-11	60.0 (9.9)	61.7 (9.4)	58.4 (10.2)	1.40 (*68*)	.166
BDI	4.2 (4.4)	4.6 (4.8)	3.8 (4.0)	0.76 (*64*)	.449
STAI *state*	46.9 (3.4)	46.6 (2.8)	47.1 (3.9)	-0.54 (*64*)	.594
NFC	60.7 (9.8)	58.8 (10.0)	62.5 (9.4)	-0.01 (*67*)	.111
UBOS	1.8 (0.8)	1.8 (0.7)	1.8 (0.9)	-1.35 (*63*)	.992

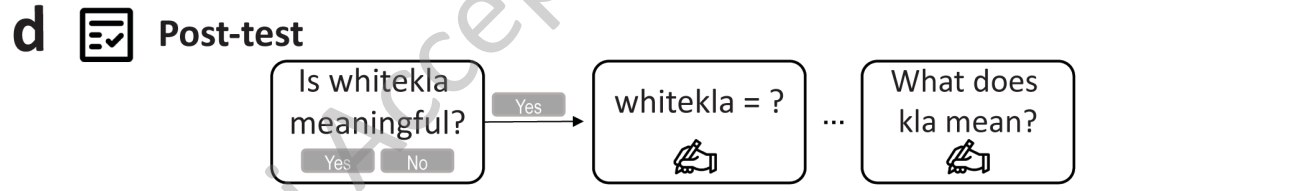
Table 1. Demographic and questionnaire scores averaged across participants and separately by treatment groups. *Mean score (SD) of demographic measures and questionnaires and the results of a two independent sample t-test on differences between treatment groups (levodopa vs. placebo). Statistical values are uncorrected for multiple comparisons. There were no significant treatment group differences after Bonferroni correction for multiple comparisons (13 comparisons threshold: p < .004). Demographics: BMI = Body Mass Index; NLV = Dutch adult reading level test (Schmand et al., 1991). Questionnaires: BIS-BAS = Behavioural Inhibition-Behavioural Activation Scale (Carver & White, 1994); BIS-11 = Barratt Impulsiveness Scale (Patton et al., 1995); BDI = Beck Depression Inventory (Beck et al., 1961); STAI = State component of the State-Trait Anxiety Inventory (Spielberger, 1983); NFC = Need For Cognition (Cacioppo et al., 1984); UBOS = Utrecht Burnout Scale (Schaufeli & van Dierendonck, 2000).*

JNeurosci Accepted Manuscript

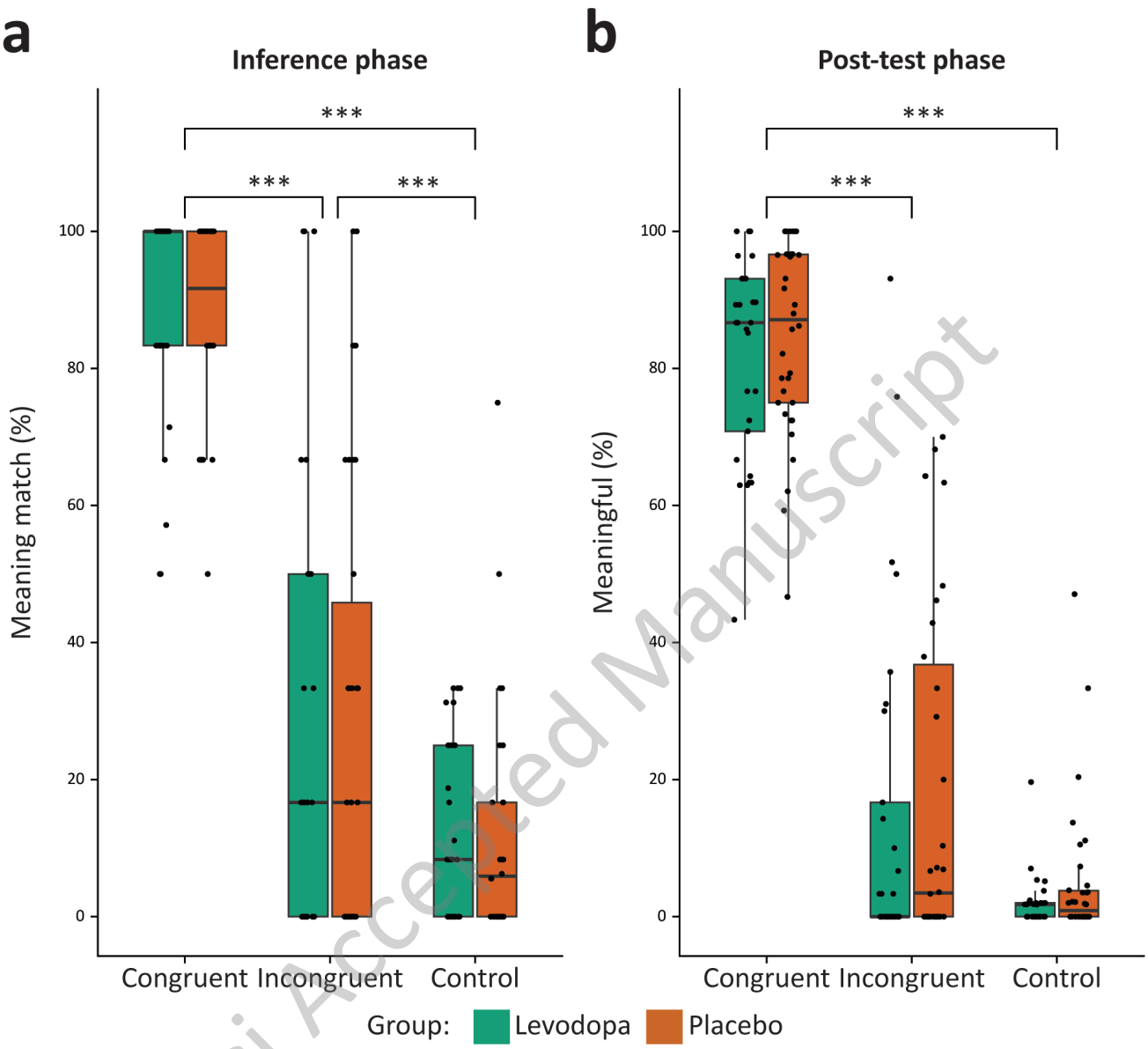




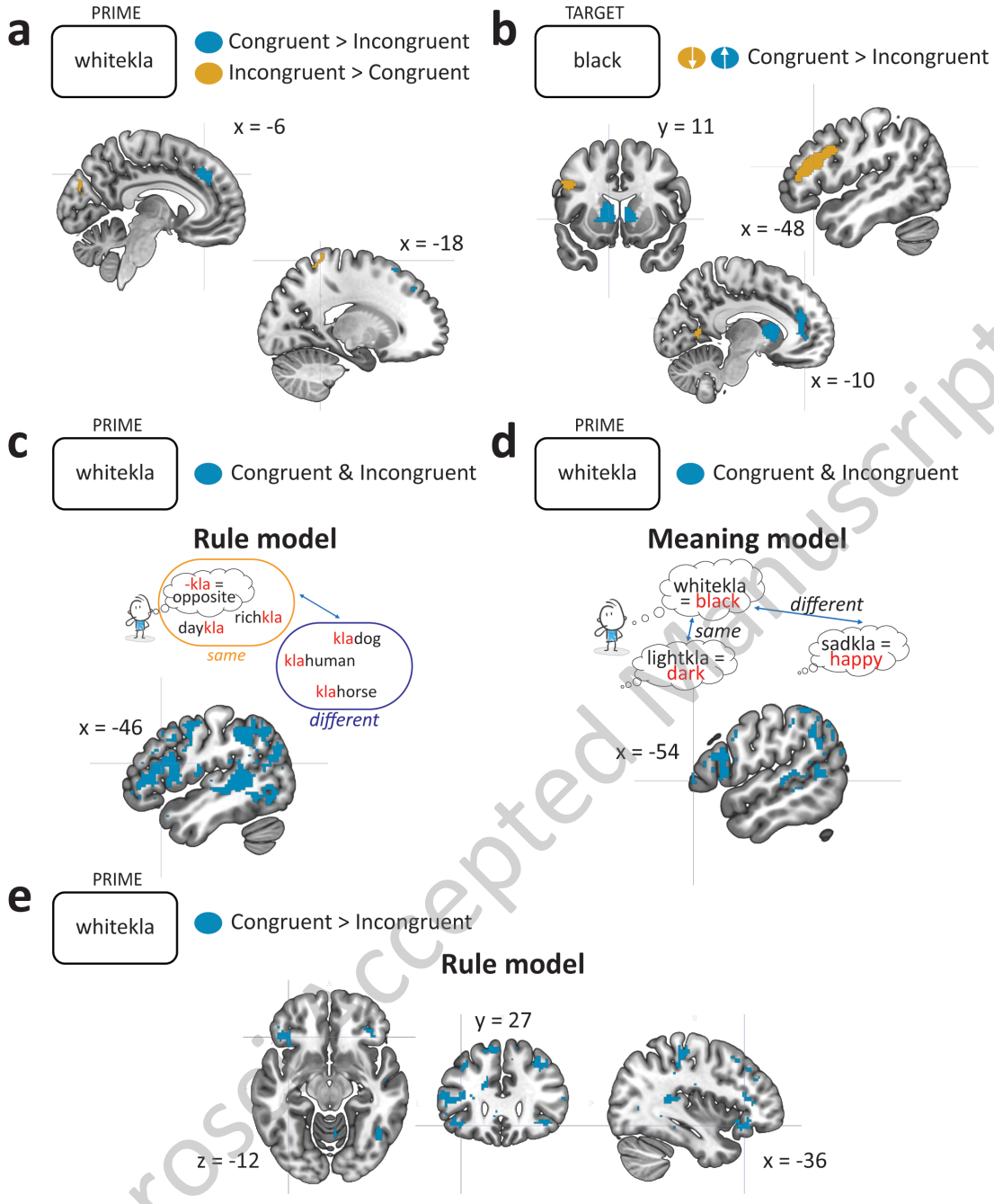




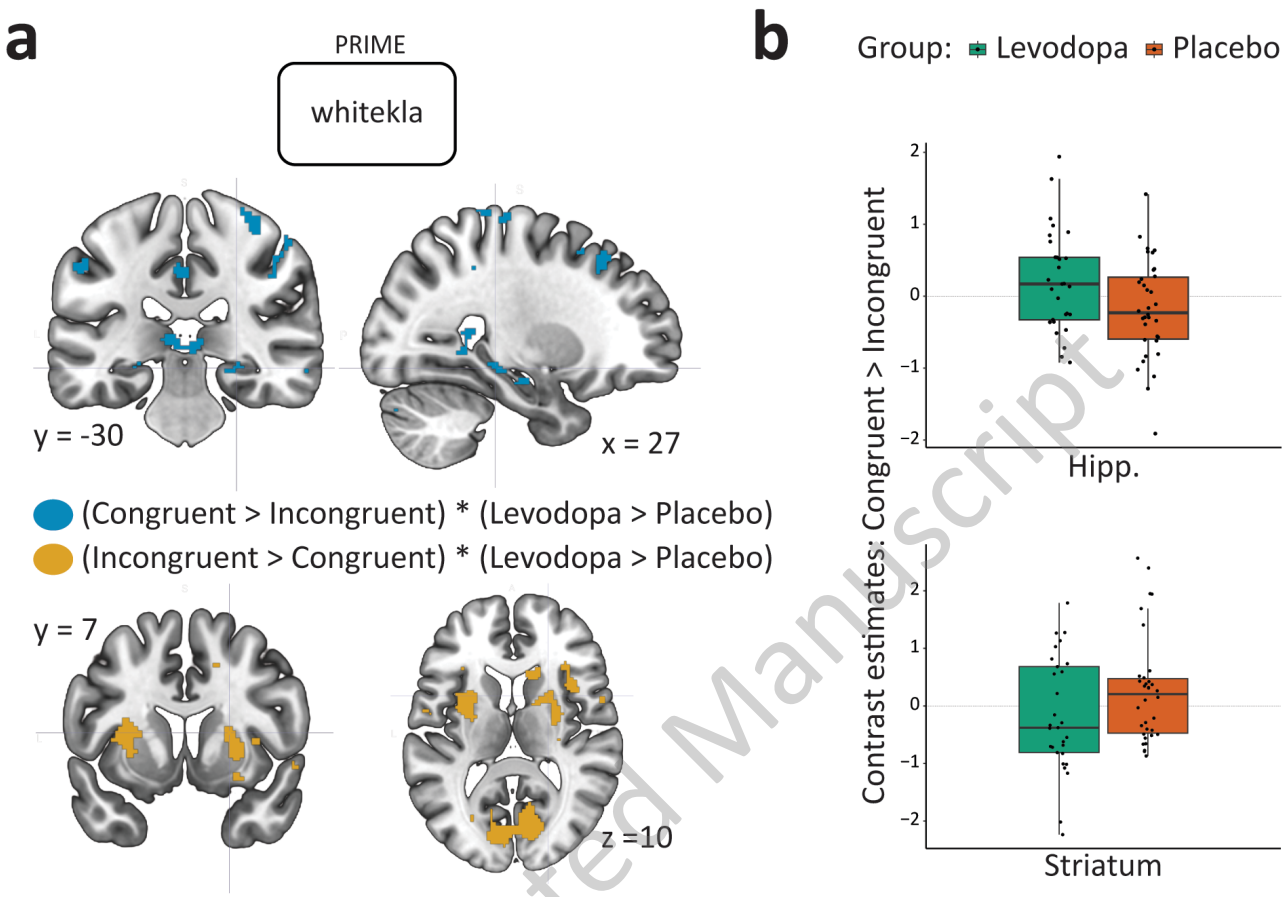
JNeurosci Accepted Manuscript



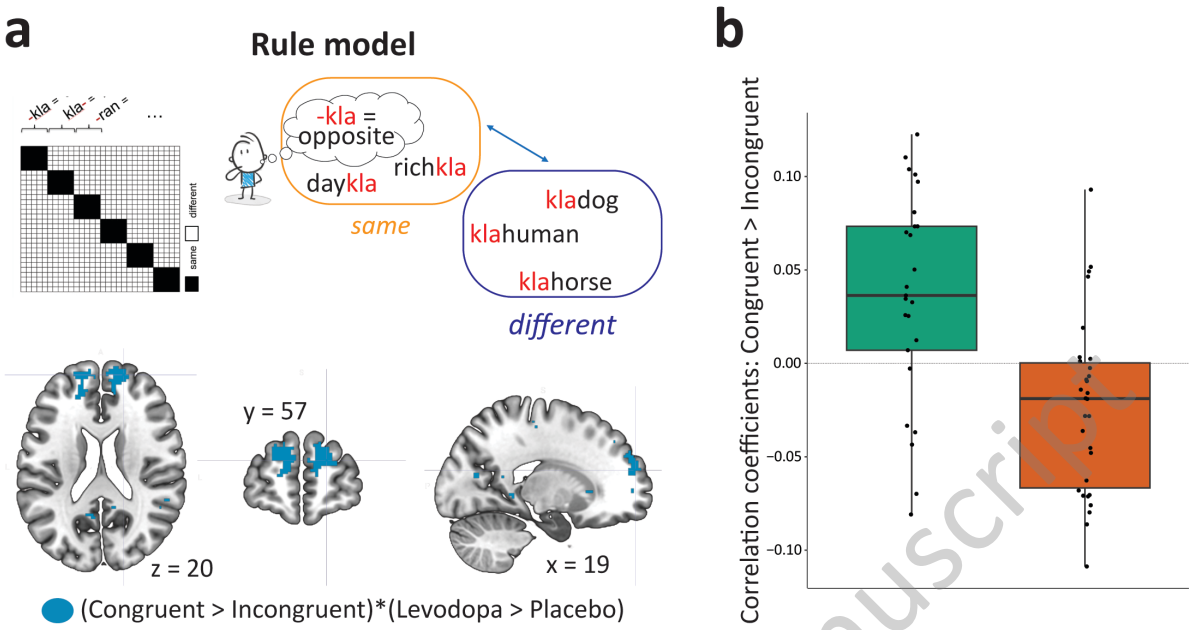
JNeurosci Accepted Manuscript

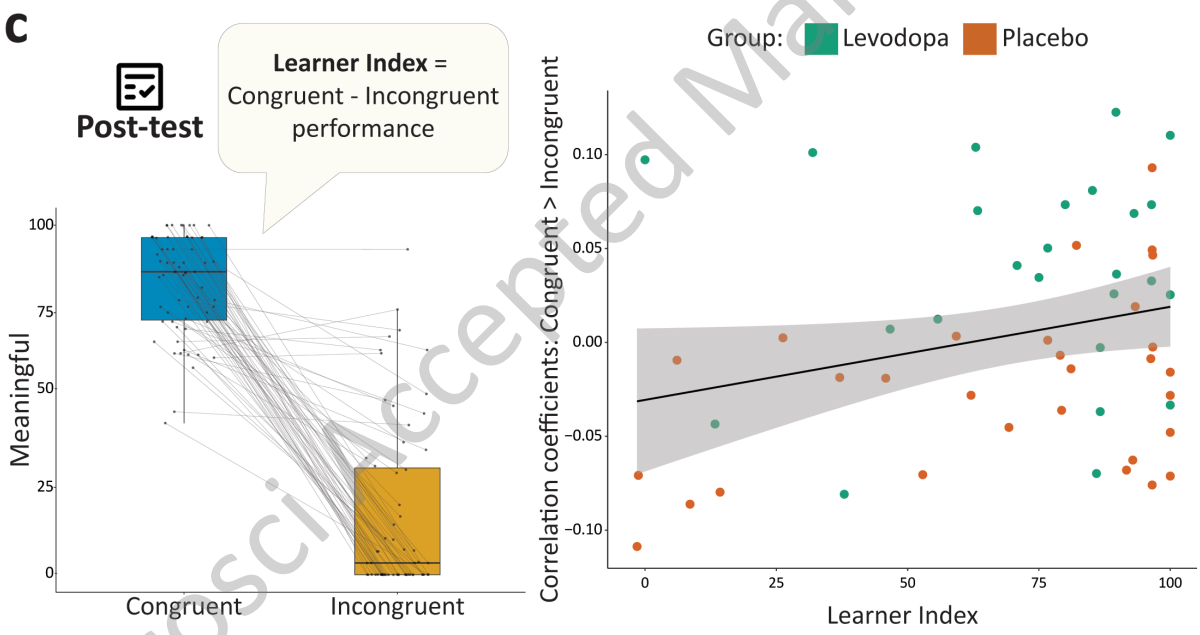


JNeurosci accepted Manuscript



JNeurosci Accepted Manuscript





JNeurosci Accepted Manuscript