Individuals often interpret the same event in different ways. How do personality traits modulate brain activity evoked by a complex stimulus? Here we report results from a naturalistic paradigm designed to draw out both neural and behavioral variation along a specific dimension of interest, namely paranoia. Participants listen to a narrative during functional MRI describing an ambiguous social scenario, written such that some individuals would find it highly suspicious, while others less so. Using inter-subject correlation analysis, we identify several brain areas that are differentially synchronized during listening between participants with high and low trait-level paranoia, including theory-of-mind regions. Follow-up analyses indicate that these regions are more active to mentalizing events in high-paranoia individuals. Analyzing participants' speech as they freely recall the narrative reveals semantic and syntactic features that also scale with paranoia. Results indicate that a personality trait can act as an intrinsic “prime,” yielding different neural and behavioral responses to the same stimulus across individuals.
hat different individuals may see the same event in dif-
ferent ways is a truism of human nature. Examples are
found at many scales, from low-level perceptual judgments
to interpretations of complex, extended scenarios. This latter
phenomenon is known as the "Rashomon effect"1 after a 1950
Japanese film in which four eyewitnesses give contradictory
accounts of a crime and its aftermath, raising the point that,
for multifaceted, emotionally charged events, there may be no single
version of the truth.

What accounts for these individual differences in interpreta-
tion? Assuming everyone has access to the same perceptual
information, personality traits may bias different individuals
toward one interpretation or another. Paranoia is one such trait,
in that individuals with strong paranoid tendencies may be more
likely to assign a nefarious interpretation to otherwise neutral
events2. While paranoia in its extreme is a hallmark symptom of
schizophrenia and other psychoses, trait-level paranoia exists as a
continuum rather than a dichotomy3,4: on a behavioral level, up to
30% of people report experiencing certain types of paranoid
thoughts (e.g., “I need to be on my guard against others”) on a
regular basis5 and trait paranoia in the population follows an
exponential, rather than bimodal, distribution6.

Few neuroimaging studies have investigated paranoia as a
continuum; the majority simply contrast healthy controls and
patients suffering from clinical delusions. However, a handful of
reports from subclinical populations describe patterns of brain
activity that scale parametrically with tendency toward paranoid
or delusional ideation. For example, it has been reported that
higher-paranoia individuals show less activity in the medial
temporal lobe during memory retrieval and less activity in the
cerebellum during sentence completion7, less activity in temporal
regions during social reflection8 and auditory oddball detection9,
but higher activity in the insula and medial prefrontal cortex
(mPFC) during self-referential processing10 and differential pat-
terns of activity in these regions as well as the amygdala while
viewing emotional pictures11.

Such highly controlled paradigms enable precise inferences
about evoked brain activity but potentially at the expense of real-
world validity. For example, brain response to social threat is
often assessed with decontextualized static photographs of unfa-
miliar faces presented rapidly in series12. Compare this to threat
detection in the real world, which involves perceiving and
interacting with both familiar and unfamiliar faces in a rich,
dynamic social context. Paranoid thoughts that eventually reach
clinical significance usually have a slow, insidious onset, involving
complex interplay between a person’s intrinsic tendencies and his
or her experiences in the world. In studying paranoia and other
trait-level individual differences, then, is important to comple-
ment highly controlled paradigms with more naturalistic stimuli.

Narrative is an attractive paradigm for several reasons. First,
narrative is an ecologically valid way to study belief formation in
action. Theories of fiction posit that readers model narratives in a
Bayesian framework in much the same way as real-world informa-
tion13, and story comprehension and theory-of-mind pro-
cesses share overlapping neural resources14. Second, a
standardized narrative stimulus provides identical input, so any
variation in interpretation reflects individuals’ intrinsic biases in
how they assign salience, learn, and form beliefs. Third, from a
neuroimaging perspective, narrative listening is a continuous,
engaging task that involves much of the brain15 and yields data
lending itself to innovative, data-driven analyses such as inter-
subject correlation (ISC)16,17.

Previous work has shown that experimenters can manipulate
patterns of brain activity during naturalistic stimuli by explicitly
instructing participants to focus on different aspects of the sti-
ulus. For example, Cooper et al. reported that activity patterns
in temporal and frontal regions varied according to whether lis-
teners were told to pay attention to action-, space-, or time-
related features of short stories18. Lahnakoski et al. showed par-
ticipants the same movie twice, asking them to adopt different
perspectives each time, and found differences in neural synchrony
depending on which perspective had been taken19. Most recently,
Yeshurun et al. presented participants with a highly ambiguous
story with at least two plausible—but very different—interpreta-
tions and used explicit primes to bias each participant toward one
interpretation or the other. Responses in higher-order brain areas,
including default mode, were more similar among participants
who had received the same prime, indicating that shared beliefs
have a powerful effect on how individuals perceive an identical
stimulus20. However, while informative, these studies have all
relied on an explicit prime or instruction; they cannot explain
why individuals often spontaneously arrive at different inter-
pretations of the same stimulus.

In this work, we use participants’ intrinsic personality traits as
an implicit prime, relating individual differences in trait paranoia
to brain activity during a naturalistic task in which participants
are faced with computationally ambiguous social circumstances. Using
an original narrative, we show that while much of the brain is
synchronized across all participants during story listening, stra-
tifying participants based on trait paranoia reveals an additional
set of regions with stereotyped activity only among high-paranoia
individuals; many of these are regions involved in theory of mind
and mentalizing. An encoding model of the task suggests that
these regions, including the temporal pole and mPFC, are parti-
cularly sensitive to “mentaling events” when the main character
is experiencing an ambiguous social interaction or explicitly
reasoning about other characters’ intentions. Finally, we measure
participants’ behavioral reactions to the narrative by analyzing
their speech as they freely recall the story and identify semantic
and syntactic features that vary dimensionally with trait paranoia.
Together, results indicate that a personality trait, in this case
paranoia, can modulate both neural and behavioral responses to a
single stimulus across individuals.

Results
Behavioral data and task performance. We created a fictional
narrative to serve as the stimulus for this study. The narrative
described a main character faced with a complex social scenario
that was deliberately ambiguous with respect to the intentions
of certain characters; it was designed such that different individuals
would interpret the events as more nefarious and others as less so.
A synopsis of the story is given in Supplementary Note 1.

Twenty two healthy participants listened to a pre-recorded audio
version of the narrative (total duration = 21:50 mins., divided into
two parts) during functional magnetic resonance image (fMRI)
scanning. Following each of the three parts, participants answered
eight challenging multiple-choice comprehension questions to
ensure they had been paying attention. Performance was very
accurate (15 of the 22 subjects answered 9/9 [100%] questions
correctly, while 5 answered 8/9 [89%] correctly and 2 answered 7/9
[78%] correctly). Self-report data indicated that subjects generally
found the narrative engaging and easy to pay attention to
(engagement rating on a scale of 1 to 5: mean = 3.8, s.d. = 0.96,
median = 4, median absolute deviation [m.a.d.] = 0.72; attention
rating: mean = 4.1, s.d. = 0.87, median = 4, m.a.d. = 0.66).

During a separate behavioral visit 1 week prior to the scan,
participants completed several self-report questionnaires and
behavioral tasks to assess personality traits and cognitive abilities
(see Fig. 1a for a schematic of the experimental protocol). Our
primary measure of interest was subscale A from the Green et al.
Paranoid Thoughts Scale21 (GPTS-A), henceforth referred to as
trait paranoia score. We administered this scale on a different day and placed it among other tasks unrelated to paranoia to minimize any priming effects or demand characteristics that might influence participants’ eventual reactions to the narrative. Possible scores on the GPTS-A range from 16 to 80; higher scores are generally observed only in clinical populations. In our healthy sample, we observed a right-skewed distribution that nonetheless had some variance (range = 16–40, mean = 20.6, s.d. = 6.3; median = 18.5, m.a.d. = 4.0; see Fig. 1b for a histogram of the distribution). This is consistent with observations from much larger sample sizes that trait paranoia follows an exponential, rather than normal, distribution in the healthy population.

**Story listening evokes widespread neural synchrony.** Our primary approach for analyzing the fMRI data was ISC, which is a model-free way to identify brain regions responding reliably to a naturalistic stimulus across subjects. In this approach, the time course from each voxel in one subject’s brain across the duration of the stimulus is correlated with the time course of the same voxel in a second subject’s brain. Voxels that show high correlations in their time courses across subjects are considered to have a stereotyped functional role in processing the stimulus. The advantage of this approach is that it does not require the investigator to have an a priori model of the task, nor to assume any fixed hemodynamic response function.

In a first-pass analysis, we calculated ISC at each voxel across the whole sample of \( n = 22 \) participants, using a recently developed statistical approach that relies on a linear mixed-effects model with crossed random effects to account for the non-independent structure of the correlation matrix. Results are shown in Fig. 2. As expected, given the audio-linguistic nature of the stimulus, ISC was highest in primary auditory cortex and language regions along the superior temporal lobe, but we also observed widespread ISC in other parts of association cortex, including frontal, parietal, midline, and temporal areas, as well as the posterior cerebellum. These results replicate previous reports that complex naturalistic stimuli induce stereotyped responses across participants in not only the relevant primary cortex but also higher-order brain regions.

Also as expected, ISC was generally lower or absent in primary motor and somatosensory cortex, although we did observe significant ISC in parts of primary visual cortex and language regions along the superior temporal lobe, but we also observed widespread ISC in other parts of association cortex, including frontal, parietal, midline, and temporal areas, as well as the posterior cerebellum. These results replicate previous reports that complex naturalistic stimuli induce stereotyped responses across participants in not only the relevant primary cortex but also higher-order brain regions.

**Paranoia modulates neural response to the narrative.** Having established that story listening evokes widespread neural synchrony across all participants, we next sought to determine whether there were brain regions whose degree of ISC was modulated by trait paranoia. Using a median split of GPTS-A scores, we stratified our sample into a low-paranoia group (GPTS-A ≤ 18, \( n = 11 \)) and a high-paranoia group (GPTS-A ≥ 19, \( n = 11 \)) and compared ISC across each of these two groups. We observed widespread ISC in the high-paranoia group, but not in the low-paranoia group, suggesting that trait paranoia modulates neural response to the narrative. This finding is consistent with previous reports that suggest that trait paranoia is associated with abnormalities in neural processing of naturalistic stimuli.

***Fig. 1*** Experimental protocol, distribution of trait paranoia scores, and inter-subject correlation (ISC) analysis. **a** Schematic of the experimental protocol. Participants came to the laboratory for an initial behavioral visit, during which they completed several computerized cognitive tasks as well as self-report psychological scales, one of which was the Green et al. Paranoid Thoughts Scale (GPTS). To minimize demand characteristics and/or priming effects, the fMRI scan visit took place approximately 1 week later. During this visit, subjects listened to an ambiguous social narrative in the scanner and then completed an extensive post-narrative battery consisting of both free-speech prompts and multiple-choice items. **b** Distribution of scores on the GPTS-A subscale across \( n = 22 \) participants, and median split used to stratify participants into low (≤18, blue) and high (≥19, orange) trait paranoia. **c** Schematic of ISC analysis. Following normalization to a standard template, the ISC of activation time courses during narrative listening was computed for each voxel (\( i \), yellow square; enlarged relative to true voxel size for visualization purposes) for each pair of subjects (\( i,j \)), resulting in a matrix of pairwise correlation coefficients (\( r \) values). These values were then compared across paranoia groups using voxelwise linear mixed-effects models with crossed random effects to account for the non-independent structure of the correlation matrix.
We then used the same linear mixed-effects model described above formulated as a two-group contrast to reveal areas that are differentially synchronized across paranoia levels.

We opted for a median split rather than using raw paranoia score as a continuous covariate because of the unique challenge of an ISC-based analysis, which, to take advantage of all the information contained in the cross-subject correlation matrix (Fig. 1c), requires any covariates to be at the subject pair level, rather than the level of individual subjects. Because trait paranoia is a single scalar value per participant, it is difficult to calculate a meaningful pairwise metric. (Median splits can also mitigate the influence of extreme values, such as the two participants with GPTS-A ≥ 38 [cf. Fig. 1b], ensuring that these do not have an outsize effect on the results.) Still, we conducted post-hoc tests to investigate continuous relationships with raw GPTS-A score whenever possible to respect the inherently continuous nature of this trait and to facilitate interpretation.

We were primarily interested in three contrasts. First, which voxels show greater ISC among pairs of high-paranoia participants vs. low-paranoia participants, or vice versa? Second and third, which voxels show greater ISC among pairs of low- or high-paranoia participants, respectively (i.e., low–low or high–high), than pairs of participants mismatched for group (i.e., high–low)? All three contrasts reveal regions whose response time courses are modulated by trait paranoia in some way. These contrasts are schematized in Fig. 1c.

Results are shown in Fig. 3. In the first contrast (Fig. 3a), several regions emerged as being more synchronized in the high-paranoia group relative to the low-paranoia group. Significant clusters were found in the left temporal pole (Talairach coordinates for center of mass: [+46.7, −10.0, −26.2]), left precuneus ([+10.8, +71.0, +35.9]), and two regions of the right mPFC (one anterior [−8.1, −46.9, +16.3] and one dorsal [−2.9, −14.8, +45.1]; Fig. 3a). Searches for these coordinates on Neurosynth, an automated fMRI result synthesizer for mapping between neural and cognitive states24, indicated that, for the left temporal pole and right anterior mPFC clusters, top meta-analysis terms included “mentalizing,” “mental states,” “intentions,” and “theory mind.” There were no regions showing a statistically significant difference in the reverse direction (low paranoia > high paranoia).

In the second contrast (Fig. 3b, cool colors), pairs of low-paranoia participants were more synchronized than pairs of inter-group participants in the left lateral occipital gyrus (center of mass: [+31.3, +86.1, +14.0]). Neurosynth: “objects,” “scene,” “encoding”), and in the third contrast (Fig. 3b, warm colors), pairs of high-paranoia participants were more synchronized than pairs of inter-group participants in the right angular gyrus ([−44.8, +57.9, +37.9], Neurosynth: “beliefs”). Interestingly, there were no voxels of statistically significant overlap between the...
second and third contrasts, indicating that no single region had a time course that was equally synchronized within groups but qualitatively different between groups. Instead, for most of the regions that emerged from the three contrasts, the relationship between trait paranoia and time course synchrony is best expressed by the Anna Karenina principle: all paranoid participants are alike; all not-paranoid participants are not paranoid in their own way (except in the lateral occipital gyrus, where it is the opposite).

As these regions were obtained via dichotomization into groups, we also conducted post-hoc tests to determine whether ISC remained sensitive to finer-grained differences in trait paranoia. We were primarily interested in two regions that emerged from the first contrast, the left temporal pole and right mPFC, since these are known from prior literature to be involved in theory of mind and mentalizing. To determine whether ISC in these regions scales monotonically with trait paranoia, we visualized the participant-by-participant ISC matrices with participants ordered by trait paranoia score (Fig. 4a, c). Visual inspection suggests a relatively continuous increase in ISC values as one moves down and to the right along the diagonal, which represents pairs of increasingly high-paranoia participants. To quantify this, we plotted each participant’s median ISC with all other participants (i.e., the median of each row of the ISC matrix) against their paranoia rank within the sample (i.e., 1–22; Fig. 4b, d). For both regions of interest (ROIs), participants with higher paranoia rank tended to have higher median ISC ($r_s = 0.71$ and $r_s = 0.63$ for the left temporal pole and right mPFC, respectively; both $p < 0.002$). We used paranoia rank rather than raw score to mitigate the influence of the two participants with extreme paranoia scores (≥38; cf. Fig. 1b).

Effects are specific to paranoia. We conducted several control analyses to rule out the possibility that the observed group differences were driven by a factor other than trait paranoia. (For all analyses in this section, we checked for both categorical and continuous relationships with paranoia; full results are reported in Table 1.) For example, if the high-paranoia participants have better overall attentional and cognitive abilities, they might simply be paying closer attention to the story, inflating ISC values but not necessarily because of selective attention to ambiguous or suspicious details. However, there were no differences between high- and low-paranoia participants on any of the cognitive tasks we administered (verbal IQ, vocabulary, fluid intelligence or working memory), making it unlikely that observed differences are due to trait-level differences in attention or cognition. As for state-level attention during the story, there was no relationship between paranoia and number of comprehension questions answered correctly, total word count during the recall task, or self-report measures of engagement and attention. We also explored potential imaging-based confounds and found that...
Fig. 4 Inter-subject correlation (ISC) scales continuously with trait paranoia. Post-hoc analyses for two regions of interest (ROIs) that emerged from the dichotomized contrast between high- and low-paranoia groups (cf. Fig. 3a): left temporal pole (top row) and right medial prefrontal cortex (PFC, bottom row). a Location of ROI (left) and participant-by-participant ISC matrix (right) for the left temporal pole. Participants are ordered by increasing trait paranoia score. Each matrix element reflects the correlation between two participants’ activation time courses in the left temporal pole during narrative listening. Higher correlations are visible as one moves to the right and down along the diagonal, representing pairs of increasingly high-paranoia individuals. b Scatter plot of paranoia rank vs. median ISC value—i.e., the median of each row of the ISC matrix in a. Each dot represents a participant. Rank correlation indicates a significant monotonic relationship between trait paranoia and median ISC in left temporal pole (rs = 0.71, p = 0.0002). c Location of ROI and participant-by-participant ISC matrix for the right medial PFC. Participants are ordered as in a. d Scatter plot of each participant’s paranoia rank vs. their median ISC value in the right medial PFC. As in b, rank correlation indicates a significant monotonic relationship between paranoia rank and median ISC (rs = 0.63, p = 0.0016).

Table 1 Trait paranoia was unrelated to potential confounding variables

| Categorical (low vs. high) | Continuous |
|----------------------------|------------|
| t p | Spearman r | p |
| Demographics | | |
| Age | | |
| 0.81 | 0.43 | 0.11 | 0.62 |
| Sex | 1.64 | 0.20 | | |
| Education (years) | | |
| −0.24 | 0.81 | 0.15 | 0.49 |
| Cognitive ability | | |
| Working memory: Letter n-back (precision) | −0.45 | 0.66 | 0.16 | 0.47 |
| Fluid intelligence: Raven’s matrices (total correct of 9 items) | 0.00 | 1.00 | −0.03 | 0.89 |
| Vocabulary: WRAT Word Reading (total correct of 42 items) | −1.42 | 0.17 | 0.31 | 0.16 |
| Verbal IQ: Penn logical reasoning test (total correct of 8 items) | 0.23 | 0.82 | −0.01 | 0.96 |
| Words of 6+ letters (free recall) | −1.03 | 0.32 | 0.04 | 0.85 |
| Words per sentence (free recall) | 0.31 | 0.76 | −0.18 | 0.43 |
| fMRI data quality | | |
| Head motion (mean FD; mm) | 0.94 | 0.36 | 0.01 | 0.96 |
| No. of frames censored | −0.70 | 0.49 | −0.08 | 0.74 |
| Average tSNR | −1.12 | 0.28 | 0.23 | 0.30 |
| Attention to stimulus | | |
| No. of comprehension questions correct | −0.31 | 0.76 | 0.08 | 0.72 |
| Total word count, free recall | 1.00 | 0.33 | −0.26 | 0.24 |
| Self-reported attention | 0.48 | 0.63 | −0.02 | 0.95 |
| Self-reported engagement | 0.89 | 0.39 | −0.10 | 0.65 |

There were no significant differences between high- and low-paranoia participants in terms of demographics, cognitive abilities, fMRI data quality, or attention to the stimulus. Categorical comparisons were carried out using Student’s t-tests between the low- and high-paranoia groups as determined by median split (degrees of freedom for all t-tests = 20). Continuous comparisons were carried out using Spearman (rank) correlation between raw paranoia score and the variable of interest. All p-values are raw (uncorrected). FD framewise displacement, tSNR temporal signal-to-noise ratio, WRAT Wide Range Achievement Test.

*Measured with chi-squared test.
paranoia was not related to amount of head motion during the scan (as measured by mean framewise displacement), number of censored frames, or temporal signal-to-noise ratio. Paranoia groups did not differ in age or sex breakdown. Thus we are reasonably confident that the observed effects are driven by true trait-level differences in paranoia between individuals.

Activity to mentalizing events scales with paranoia. Results of the first contrast from the two-group ISC analysis indicated that certain brain regions showed a more stereotyped response in high-paranoia vs. low-paranoia individuals. What features of the narrative were driving activity in these regions? In theory, ISC allows for reverse correlation, in which peaks of activation in a given region’s time course are used to recover the stimulus events that evoked them. In practice, this is often difficult. Especially with narrative stimuli, in which structure is built up over relatively long timescales, it is challenging to pinpoint exactly which event—word, phrase, sentence—triggered an increase in BOLD activity.

Rather than rely on reverse correlation, a data-driven decoding approach, we took an encoding approach: we modeled events in the task that we hypothesized would stimulate differing interpretations across individuals and evaluated the degree to which certain ROIs responded to such events, using a general linear model (GLM) analysis. Specifically, we labeled sentences in the story when the main character was experiencing an ambiguous (i.e., possibly suspicious) social interaction and/or sentences when she was explicitly reasoning about the intentions of other characters. For brevity, we refer to these time points as “mentalizing events.” In creating the regressor, all events were time-locked to the end of the last word of the labeled sentences, when participants are presumably evaluating information they just heard and integrating it into their situation model of the story.

We hypothesized that the two ROIs from the previous analysis known to be involved in theory of mind and mentalizing, the left temporal pole and right mPFC, would be more active to mentalizing events in individuals with higher trait paranoia. We included two additional ROIs, the left tempo-parietal junction (TPJ) and left Heschl’s gyrus, as a positive and negative control, respectively. We selected the left TPJ as a positive control because of its well-established role in theory-of-mind and mentalizing processes and the fact that it emerged as highly synchronized across all participants (cf. Fig. 2) but did not show a group difference (cf. Fig. 3); thus we hypothesized that this region should respond to mentalizing events in all participants, regardless of trait paranoia. Conversely, left Heschl’s gyrus (primary auditory cortex) should only respond to low-level acoustic properties of the stimulus and not show preferential activation to mentalizing events in either group or the sample as a whole. See Fig. 5a for ROI locations.

For each participant, we regressed the time course of each of these four ROIs against the mentalizing-events regressor and compared the resulting regression coefficients between groups (Fig. 5b). Compared to low-paranoia individuals, high-paranoia individuals showed stronger responses in both the left temporal pole (two-sample t(20) = 2.71, padj = 0.014) and right mPFC (t(20) = 3.36, padj = 0.007). As hypothesized, responses in the left TPJ were strong across the whole sample (one-sample t(21) =...
8.73, \( p < 0.0001 \)), but there was no significant difference between groups in this ROI (\( t(20) = 0.67, p_{\text{adj}} = 0.34 \)). Also as hypothesized, the sample as a whole did not show a significant response to these events in primary auditory cortex (one-sample \( t(21) = 0.44, p = 0.66 \)), and there was no group difference (\( t(20) = 0.47, p_{\text{adj}} = 0.34 \)).

To confirm that these results hold if paranoia is treated as a continuous variable, we conducted additional post-hoc tests in which we correlated participants’ paranoia ranks and regression coefficients for all four ROIs (Fig. 5d). As expected, response to suspicious events was significantly related to paranoia rank in the left temporal pole (\( r = 0.57, p = 0.005 \)) and right mPFC (\( r = 0.64, p = 0.001 \)) but not in the left TPI (\( r = -0.04, p = 0.86 \)) or left Heschl’s gyrus (\( r = 0.02, p = 0.95 \)).

As an additional control, to check if this effect was specific to mentalizing events and not just any sentence offset, we created an inverse regressor comprising all non-mentalizing events (i.e., by flipping the binary labels from the mentalizing-events regressor, such that all sentences were labeled except those containing an ambiguous social interaction or explicit mentalizing as described above). There were no differences between paranoia groups in any of the four ROIs in response to non-mentalizing sentences (Fig. 5c) and no continuous relationships between regression coefficient and paranoia rank (Fig. 5c). This indicates that trait paranoia is associated with differential sensitivity of the left temporal pole and right mPFC to not just any information but specifically to socially ambiguous information that presumably triggers theory-of-mind processes.

Paranoia modulates behavioral response to the narrative.

Having established that trait paranoia modulates individuals’ brain responses to an ambiguous narrative, we next sought to determine whether this trait also modulates their behavioral responses to the narrative. In other words, does trait-related (intrinsic) paranoia bear upon state-related (stimulus-evoked) paranoia? If the observed differences in neural activity propagate up to conscious perception and interpretation of the stimulus, then participants’ subjective experiences of the narrative should also bear a signature of trait paranoia.

Immediately following the scan, participants completed a post-narrative battery that consisted of free-speech prompts followed by multiple-choice items to characterize their beliefs and feelings about the story. For the first item, participants were asked to retell the story in as much detail as they could remember, and their speech was recorded. Participants were allowed to speak for as long as they wished on whatever aspects of the story they chose. Without guidance from the experimenter, participants recalled the story in rich detail, speaking an average of 1081 words (range = 399–3185, s.d. = 610).

Audio recordings of participants’ speech were transcribed and submitted to the language analysis software Linguistic Inquiry and Word Count (LIWC). The output of LIWC is one vector per participant describing the percentage of speech falling into various semantic and syntactic categories. Example semantic categories are positive emotion (“love,” “nice”), money (“cash,” “owe”), and body (“hands,” “face”), while syntactic categories correspond to parts of speech such as pronouns, adjectives, and prepositions; there are 67 categories in total.

Using partial least-squares regression, we searched for relationships between speech features and trait paranoia score. More than 72% of the variance in trait paranoia score could be accounted for by the first component of speech features; the loadings of semantic and syntactic categories for this component are visualized in Fig. 6a. The feature with the highest positive loading—indicating a positive relationship with paranoia—was affiliation, a category of words describing social and familial relationships (e.g., “ally,” “friend,” “social”). Also associated with high trait paranoia was frequent use of adjectives as well as anxiety- and risk-related words (e.g., “bad,” “crisis”); drives, a meta-category that includes words concerning affiliation, achievement, power, reward, and risk; and health-related words (e.g., “clinic,” “fever,” “infected”); recall that the story featured a doctor treating patients in a remote village; cf. Supplementary Note 1). Features with strongly negative loadings—indicating an inverse relationship with paranoia—included male references (e.g., “him,” “his,” “man,” “father”); anger-related words (“yell,” “annoyed”); function words (“it,” “from,” “so,” “with”); and conjunctions (“and,” “but,” “until”). Figure 6b contains specific examples for selected categories from participants’ speech transcripts.

After the free-speech prompts, participants answered a series of multiple-choice questions (see Supplementary Table 1 for the full questionnaire). First, they were asked to rate the degree to which they were experiencing various emotions (suspicion, paranoia, sadness, happiness, confusion, anxiety, etc; 16 in total) on a scale from 1 to 5. Most of ratings skewed low—for example, the highest paranoia rating was 3, and only six subjects rated their paranoia level >1. Interestingly, there was no significant correlation between trait paranoia score and self-reported paranoia (\( r = -0.02, p = 0.91 \)) or suspicion (\( r = -0.11, p = 0.62 \)) following the story. Neither were any of the other emotion ratings significantly correlated with trait-level paranoia (all uncorrected \( p > 0.12 \); see Fig. 6c).

Second, participants were asked to rate the three central characters on six personality dimensions (trustworthy, impulsive, considerate, intelligent, likeable, naive; see Supplementary Fig. 1a). Third, they were asked to rate the likelihood of each of the six scenarios (see Supplementary Fig. 1b), and finally, to indicate (via forced-choice options) what they believed the main character would do next, as well as what they themselves would do in her situation.

None of the individual questionnaire items significantly correlated with trait paranoia. However, to facilitate comparison with the speech data, we submitted the questionnaire data to a second partial least-squares regression to search for multidimensional relationships. This analysis revealed a first component of questionnaire responses that accounted for 62% of the variance in trait paranoia (Supplementary Fig. 1c). Features with the highest positive loadings, indicating a positive relationship with paranoia, included certain answers about what individuals thought the main character might do next as well as what they would do in her place (e.g., escape from the situation), as well as feeling more uncomfortable and suspicious following the story. Features with the highest negative loadings, indicating an inverse relationship with paranoia, included feeling more amused, inspired, and hopeful following the story, as well a tendency to agree with one of the scenarios (“Juan and the other villagers had not known anything about the disease before Carmen arrived”).

Overall, then, we found signatures of paranoia in story-evoked behavior using both free-speech and self-report measures. Participants’ free speech was slightly more sensitive than their answers on the multiple-choice questionnaire. Self-report is a coarse measure that may suffer from response bias; behavior provides a richer feature set that allows for the discovery of more subtle associations. In studying nuanced individual differences, then, these results highlight the desirability of capturing behavior in both traditional and naturalistic ways.

Discussion

Here we have shown that a personality trait can act as a lens, or “implicit prime,” through which individuals perceive ambiguous
events, shaping both their neural and behavioral responses to an identical stimulus. Previous work using naturalistic tasks has shown that brain activity and behavioral responses are sensitive to experimenter instructions, i.e., an explicit prime\textsuperscript{19,20} or to the nature of the stimulus itself, i.e., whether it is more or less compelling or entertaining\textsuperscript{26–28}. The present study extends these results in an important new direction, suggesting that there is substantial implicit variation in the brain’s response to a naturalistic stimulus that stems from trait-level individual differences.

Our results have implications for the neural correlates of both trait- and state-related paranoia. Those with higher trait paranoia may have more stereotyped brain responses because suspicious and/or paranoid schemas come to mind more readily for these individuals; the idea that certain individuals tend to engage certain constructs more frequently across time and situations has been termed “chronic accessibility”\textsuperscript{29}. The relative hyperactivity of theory-of-mind regions to mentalizing events in high-paranoia individuals fits with the conception of paranoia as “over-menta-
lizing” or the tendency to excessively attribute (malevolent) intentions to other people’s actions\textsuperscript{30}. Both regions of differential response, the temporal pole and mPFC, are sometimes, but not always, reported in theory-of-mind tasks broadly construed\textsuperscript{31}; individual differences may at least partially explain the inconsistencies in the literature.

While the present study included only healthy controls with subclinical paranoia, it may provide a useful starting point for the study of paranoid or persecutory delusions in schizophrenia and related illnesses. Delusions with a persecutory theme account for roughly 70–80% of all delusions. This high prevalence is stable across time\textsuperscript{32} and geo-cultural factors\textsuperscript{33–36}, suggesting a strong biological component. Persecutory delusions are also the type most strongly associated with anger and most likely to be acted
upon, especially in a violent manner. Therefore understanding the neurobiological basis of paranoid delusions is a critical problem in psychiatry. But because delusions typically have a slow, insidious onset, it is nearly impossible to retrospectively recover triggering events in individual patients. A related challenge is that, while thematically similar, each patient’s delusion is unique in its details. Thus it is difficult to devise material that will evoke comparable responses across patients. One solution is to craft a model context using a stimulus that is ambiguous yet controlled—i.e., identical across participants, permitting meaningful comparisons of time-locked evoked activity—such as the one used in this work. Paradigms such as this one may shed light on mechanisms of delusion formation and/or provide eventual diagnostic or prognostic value.

While there is little work investigating brain activity during naturalistic stimuli in psychiatric populations, a handful of studies have used such paradigms in autism, finding that autistic individuals are less synchronized with one another and with typically developing controls while watching movies of social interactions. The degree of asynchrony scales with autistic spectrum phenotype severity in both the autistic and control groups. It is interesting to juxtapose these reports with the present results, in which individuals with a stronger paranoia phenotype were more synchronized during exposure to socially relevant material; ultimately, this fits with the notion of autism and psychosis as opposite ends of the same spectrum, involving hypo- and hyper-mentalization, respectively. Future studies should combine naturalistic stimuli with ISC-based analyses that cut across diagnostic labels to examine how neural responses vary across the full range of human phenotypes.

From a methodological perspective, much of the fMRI research on individual differences has shifted in recent years from measuring activation in task-based conditions to measuring functional connectivity, predominantly at rest. Both paradigms suffer from limitations: traditional tasks are so tightly controlled that they often lack ecological validity; resting-state scans, on the other hand, are entirely unconstrained, making it difficult to separate signal from noise. Naturalistic tasks may be a happy medium for studying both group-level functional brain organization as well as individual differences. We and others argue that such tasks could serve as a “stress test” to draw out individual variation in the brain and behaviors of interest, enhancing signal in the search for neuroimaging-based biomarkers and permitting more precise inferences about the sources of individual differences in neural activity.

**Methods**

**Participants.** A total of 23 healthy volunteers participated in this study. Data from one participant were excluded owing to excessive head motion and self-reported falling asleep during the last third of the narrative. Thus, the final data set used for analysis consisted of 22 participants (11 females; age range = 19–35 years, mean = 27, s.d. = 4.4). All participants were right-handed, native speakers of English, with no history of neurological disease or injury, and were not on psychoactive medication at the time of scanning. All participants provided written informed consent in accordance with the Institutional Review Board of Yale University. The experiment took place over two visits to the laboratory. Participants were paid $25 upon completion of the first visit (behavioral assessments) and $75 upon completion of the second visit (fMRI scan); all participants completed both visits.

**Stimulus.** An original narrative was written by author E.S.F. to serve as the stimulus for this experiment. For a synopsis of the story, see Supplementary Note 1. The full audio recording, as well as a complete transcript, are available in the “stimuli” directory at the following URL: https://openneuro.org/datasets/d001338/. To mitigate confounds associated with education level or verbal IQ, we wrote the narrative text to be easy to comprehend, with a readability level of 78.1/100 and a grade 5.5 reading level as calculated by the Flesch–Kincaid Formula.

A male native speaker of English read the story aloud and his speech was recorded using high-quality equipment at Haskins Laboratories (New Haven, CT). The speaker was instructed to read in a natural, conversational tone, but without excess emotion. The final length of the audio recording was 21:50.

**Experimental protocol.** The experiment took place over two visits to the laboratory. Visit 1 was purely behavioral and took place approximately 1 week prior to visit 2 (MRI scan). During visit 1, participants completed a battery of self-report and behavioral tasks. While our primary measure of interest was the GPTS, we also administered several other psychological scales and cognitive assessments, in part to help reduce any demand characteristics that would allow participants to intuit the purpose of the study. We chose the GPTS because it provides a meaningful assessment of trait-level paranoia in clinical, but crucially, also in subclinical and healthy populations. In a previous study, score on this scale best predicted feelings of persecution following immersion in a virtual-reality environment. The full GPTS contains two subscales, A and B, which pertain to ideas of social referentiality and ideas of persecution, respectively. We focused on subscale A, as it produces a wider range of scores in subclinical populations.

The following cognitive tests were administered via the web interface of the University of Pennsylvania Computerized Neuropsychological Test Battery (PennCNP; http://penncnp.med.upenn.edu/). SRAVEN (short Raven’s progressive matrices), a measure of abstraction and mental fluid intelligence; SPVRT (short Penn logical reasoning test, a measure of verbal intelligence); and LBRE (letter–back, a measure of working memory). We also administered the word reading test from the Wide Range Achievement Test (WRAT), a measure of reading and vocabulary.

**MRI data acquisition and preprocessing.** Scans were performed on a 3 T Siemens TimTrio system at the Yale Magnetic Resonance Research Center. After an initial localization scan, a high-resolution three-dimensional volume was collected using a magnetization-prepared rapid gradient echo sequence (208 contiguous sagittal slices, slice thickness = 1 mm, matrix size 256 × 256, field of view = 256 mm, TR = 2400 ms, TE = 1.9 ms, flip angle = 8°). Functional images were acquired using multiband T2*-sensitive gradient-recalled single-shot echo-planar imaging pulse sequence (TR = 1000 ms, TE = 30 ms, voxel size = 2.0 mm3, flip angle = 60°, resolution = 3.0 × 3.0 × 3.0 mm3). The subject was instructed to think of nothing in particular; (2) a movie-watching run using Inscapes (7:00/420 TRs); (3) three narrative-listening runs corresponding to parts 1, 2, and 3 of the MRI scan. The full audio recording was divided into three segments of length 8:46, 7:32, and 5:32, respectively; each of these segments was delivered in a continuous functional run while participants were in the scanner. To ensure attention, after each run, subjects answered three challenging multiple-choice comprehension questions regarding the content of the part they had just heard, for a total of nine questions. Immediately after each scan, participants completed a post-narrative questionnaire that consisted of open-ended prompts to elicit free speech, followed by multiple-choice items. These are described further below.

We acquired the following functional scans: (1) an initial eyes-open resting-state run (6:00/360 TRs in duration) during which subjects were instructed to relax and think of nothing in particular; (2) a movie-watching run using Inscapes (7:00/420 TRs); (3) three narrative-listening runs corresponding to part 1 of the MRI scan; (4) a post-narrative, eyes-open resting-state run (6:00/360 TRs) during which subjects were instructed to reflect on the story they had just heard. The present work focuses exclusively on data acquired during narrative listening. The narrative stimulus was delivered through MRI-compatible audio headphones and a short “volume check” scan that was conducted prior to the first narrative run to ensure that participants could adequately hear the stimulus above the scanner noise. To promote engagement, during the three narrative runs, participants were asked to fixate on a static image of a jungle settlement and to actively imagine the story events as they unfolded.

Following conversion of the original DICOM images to NIFTI format, AFNI (Cox 1996) was used to preprocess MRI data. The functional time series went through the following preprocessing steps: despiking, head motion correction, affine alignment with anatomy, nonlinear alignment to a Talairach template (TT_N27), and smoothing with an isotropic full-width half-maximum of 5 mm. A voice mask was defined on the template and intersected with the subject’s cerebellar fluid mask to make a subject-specific ventricle mask. Regressors were created from the first three principal components of the ventricles, and fast ANATCOR (Jo et al. 2010) was implemented to provide local white matter regressors. Additionally, the subject’s six motion time series, their derivatives, and linear polynomial baselines for each of the functional runs were included as regressors. Censoring of time points was performed whenever the per-time motion (Euclidean norm of the motion derivatives) was ≥0.3 or when ≥10% of the brain voxels were outliers. Censored time points were set to zero rather than removed altogether (this is the conventional way to do censoring, but especially important for fMRI analysis, to preserve the temporal structure across participants). The final output of this preprocessing pipeline was a single functional run concatenating data from the three story runs (total duration = 21:50, 1310 TRs). All analyses were conducted in volume space and projected to the surface for visualization purposes.

We used mean framewise displacement (MFD), a per-participant summary measure to assess the amount of head motion in the sample. MFD was overall relatively low (after censoring: mean = 0.075 mm, s.d. = 0.026, range = 0.035–0.14). Number of censored time points during the story was overall low but...
followed a right-skewed distribution (range = 0–135, median = 4, median absolute deviation = 25). All 22 participants in the final analysis retained at least 89% of the total time points in the story, so missing data was not a substantial concern. Still, we performed additional control analyses to ensure that number of censored time points and amount of head motion were not associated with paranoia score in any way that would confound interpretation of the results (see Table 1).

**Inter-subject correlation.** Following preprocessing, ISC during the story was computed across all possible pairs of subjects \((s_j)\) using AFNI’s 3dCorrLag function, resulting in 231 \((m^2−n^2)/2\), where \(m = 22\) is the unique ISC value where the value at each voxel represents the Pearson’s correlation between that voxel’s time course in subject \(i\) and its time course in subject \(j\).

To identify voxels demonstrating statistically significant ISC across all 231 subject pairs, we performed inference at the single-group level using a recently developed linear mixed-effects (LME) model with a crossed random-effects formulation to accurately account for the correlation structure embedded in the ISC data\(^{22}\). This approach has been characterized extensively, including a comparison to non-parametric approaches, and found to demonstrate proper control for false positives and good power attainment\(^{27}\). The resulting map was corrected for multiple comparisons and then clustered for visualization using a voxelwise false discovery rate threshold of \(q = 0.001\) (Fig. 2).

In a second analysis, we stratified participants according to a median split of scores on the GPTS-A subscale. We used these groups to identify voxels that had higher ISC values within one paranoia group or the other or higher ISC values within than across paranoia groups. To this end, we used a two-way formulation of the LME model. This model gives the following outputs: voxelwise population ISC values within group 1 (G1); voxelwise population ISC values within group 2 (G2); voxelwise population ISC values between the two groups that reflect the ISC effect between any pair of subjects with each belonging to different groups (G12). These outputs can be compared to obtain several possible contrasts.

For multiple contrasts and three cluster corrections using a voxelwise discovery rate threshold of \(q = 0.001\) (Fig. 2).

**Event-related analysis.** A forced-aligner (Gentle; https://lowerquality.com/gentle/) was used to obtain precise timing information for each word in the narrative, by aligning the audio file with its transcript. One of the authors (E.S.F.) manually labeled sentences containing either an ambiguous social interaction or an instance of the GPTS-A item: “monitoring about others’ intentions using a binary scoring system (1 = ambiguous social interaction or mentalizing present in sentence, 0 = neither ambiguous social interaction nor mentalizing present). Four additional, independent raters previously naive to the narrative listened to the same audio files and labeled the same sentences. Words in the semantic category “leisure” accounted for 0–1.09%. To give approximately equal weight to all categories, we standardized the data (to have zero mean unit variance) across participants before performing partial least squares regression (PLSR) as described in the next section. This ensures that the resulting PLS components are not simply dominated by variance in categories that are represented heavily in human speech.

**Relating story-evoked behavior to paranoia.** To determine which speech features were most related to trait paranoia, we submitted the data to a PLSR with the z-scored speech features as \(X\) (predictors) and trait paranoia score as \(Y\) (response), implemented in Matlab as plregress. PLSR is a latent variable approach to modeling between two matrices, which seeks to find the direction in \(X\) space that explains the maximum variance in \(Y\) space. It is well suited to the current problem, because it can handle a predictor matrix with more variables than observations, as well as multi-collinearity among the predictors.

In a first-pass analysis, we ran a model with 10 components to determine the number of components needed to explain most of the variance in trait paranoia. Results of this analysis indicated that the first component was sufficient to explain 72.3% of the total variance in paranoia score, so we selected just this component for visualization and interpretation. Feature loadings for this component are visualized in Fig. 6a.

**Multiple-choice questionnaires.** Following the free-speech prompts, we had participants complete a multiple-choice questionnaire to assess their feelings toward and beliefs about the story. A full list of items is provided in Supplementary Table 1; there were 47 in total.

**Analysis of speech features.** Audio recordings of participants’ retelling of the story were professionally transcribed by a third-party company. We submitted the resulting transcripts to LIWC (www.liwc.net)\(^{25}\), a software program that takes as input a given text and counts the percentage of words falling into different syntactic and semantic categories. Because LIWC was developed by researchers with interests in social, clinical, health, and cognitive psychology, the language categories were created to capture people’s social and psychological states.

We restricted LIWC output to the 67 linguistic (syntactic and semantic) categories, excluding categories relating to metadata (e.g., percentage of words found in the LIWC dictionary), as well as categories irrelevant to spoken language (e.g., punctuation). Thus our final LIWC output was a 22 × 67 matrix where each row corresponds to a participant and each column to a category.

These categories can be scaled very differently from one another. For example, words in the syntactic category “pronoun” accounted for between 10.3 and 20.5% of speech transcripts, while words in the semantic category “leisure” accounted for only 0–1.09%. To give approximately equal weight to all categories, we standardized each category (to have zero mean unit variance) across participants before performing partial least squares regression (PLSR) as described in the next section. This ensures that the resulting PLS components are not simply dominated by variance in categories that are represented heavily in human speech.

**Code availability.** More information about this project, including links to code and other supporting material, can be found at: https://esfnn.github.io/projects/ParanoiaStory.html.

**Data availability.** Source data generated during this study, including raw MRI data and the full narrative stimulus (audio and text), are available at: https://openneuro.org/datasets/ds001338/.
59. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. USA 113, 7900–7905 (2016).

60. Zwaan, R. A. & Radvansky, G. A. Situation models in language comprehension and memory. Psychol. Bull. 123, 162–185 (1998).

61. Johnson-Laird, P. N. Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness. Cambridge, MA: Harvard University Press (1983).

62. Van Dijk, T. A., Kintsch, W. & Van Dijk, T. A. Strategies of Discourse Comprehension. (Academic Press, New York, NY, 1983).

63. Whitney, C. et al. Neural correlates of narrative shifts during auditory story comprehension. Neuroimage 47, 360–366 (2009).

64. Zacks, J. M. et al. Human brain activity time-locked to perceptual event boundaries. Nat. Neurosci. 4, 631 (2001).

65. Schönwiesner, M. et al. Heschl’s gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. J. Neurophysiol. 97, 2075–2082 (2007).

Acknowledgements

This research was supported by a National Science Foundation Graduate Research Fellowship to E.S.F. and by the Intramural Research Program of the NIMH (annual report ZIAMH002783). Portions of this study used the computational capabilities of the NIH HPC Biowulf cluster (hpc.nih.gov). The authors thank Paul Taylor for advice on preprocessing pipelines for naturalistic task data, Richard Betzel for sharing code to visualize component feature loadings (Fig. 6a and Supplementary Fig. 2), and Javier Gonzalez-Castillo, Daniel Handwerker, David Jangraw, Peter Molfese, Monica Rosenberg, and Tamara Vanderwal for helpful discussion.

Author contributions

E.S.F., P.R.C., and R.T.C conceived the study. E.S.F. developed experimental materials and performed data collection. E.S.F. and G.C. analyzed the data. E.S.F., P.R.C., P.A.B., and R.T.C interpreted results. E.S.F. wrote the manuscript with comments from all other authors.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04387-2.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018