A Tradeoff in the Neural Code across Regions and Species

HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2019 October 09.

Published in final edited form as:

Cell. 2019 January 24; 176(3): 597–609.e18. doi:10.1016/j.cell.2018.12.032.

A tradeoff in the neural code across regions and species

Raviv Pryluk[1], **Yoav Kfir**[1], **Hagar Gelbard-Sagiv**[2], **Itzhak Fried**[3,4,5], **Rony Paz**[1,6,*]

[1]Department of Neurobiology, Weizmann Institute of Science, Israel

[2]Department of Physiology & Pharmacology, Sackler School of Medicine, Tel Aviv University, Israel

[3]Department of Neurosurgery, University of California Los Angeles, Los Angeles, USA

[4]Department of Neurology and Neurosurgery, Sackler Faculty of Medicine, Tel Aviv University, Israel

[5]Functional Neurosurgery Unit, Tel-Aviv Medical Center, Tel-Aviv, Israel

[6]Lead Contact

Summary

Many evolutionary years separate humans and macaques, and whereas the amygdala and cingulate-cortex evolved to enable emotion and cognition in both, an evident functional gap exists. Although it was traditionally attributed to differential neuroanatomy, functional differences might also arise from coding mechanisms. Here, we find that human neurons better utilize information capacity (efficient coding) than macaque neurons, in both regions; and that cingulate neurons are more efficient than amygdala neurons, in both species. In contrast, we find more overlap in the neural vocabulary and more synchronized activity (robustness coding) in monkeys in both regions, and in the amygdala of both species. Our findings demonstrate a tradeoff between robustness and efficiency across species and regions. We suggest that this tradeoff can contribute to differential cognitive functions between species, and underlie the complementary roles of the amygdala and the cingulate-cortex. In turn, it can contribute to fragility underlying human psychopathologies.

Introduction

The primate brain enables complex cognitive and emotional processes, yet is vulnerable to psychopathologies such as anxiety and mood-disorders that are tightly related to the balance between the amygdala and the cingulate-cortex (Averbeck and Chafee, 2016; Etkin et al., 2016; Herry and Johansen, 2014; Likhtik and Paz, 2015; Quirk and Beer, 2006; Salzman and Fusi, 2010). The amygdala underlies survival skills and their modern-day analogues – emotions, emotional learning, and social-behaviors (Adolphs, 2010; Duvarci and Pare, 2014;

[*]**Correspondence** should be addressed to R.P. (rony.paz@weizmann.ac.il).

Author Contribution

R. Pryluk. and R. Paz designed and performed the study. Y.K., H.G.S and I.F. performed the experiments. R.Pryluk analyzed the data and developed the methods. Y.K., H.G.S and I.F. contributed to data analyses and editing of the manuscript. R.Pryluk and R. Paz wrote the manuscript.

Declaration of Interests: The authors declare no competing interests.

Murray, 2007; Phelps and LeDoux, 2005), and the cingulate-cortex is involved in cognitive processes as motivation, decision making, error monitoring, and flexible adaptive learning (Heilbronner and Hayden, 2016; Kolling et al., 2016; Shenhav et al., 2016). Both regions evolved extensively in primates and form a dense reciprocal synaptic network (Barton and Aggleton, 2000; Ghashghaei et al., 2007).

Most approaches to the evolutionary differences emerging across species and brain regions highlight neuroanatomical differences such as the large size of the human brain relative to the body (MacLeod et al., 2003), and specifically the size of the neocortex across mammalian species (Barrickman et al., 2008; Byrne and Corp, 2004; MacLean et al., 2014). Recently, it has been shown that the number of neurons can vary greatly across brains of nearly the same size (Herculano-Houzel, 2016), suggesting that the number of neurons is a major factor (Gabi et al., 2016). Together, it was proposed that the large number of neurons in an increased number of cortical areas enable high cognitive capabilities (Kaas and Herculano-Houzel, 2017). Differential neuroanatomy was recognized in David Marr's classical terminology as one level of an information processing device - the hardware implementation (Marr, 1982), yet functional differences might arise from the representational and computational levels as well (Chater et al., 2006; Marr, 1982). Despite this, these levels are rarely compared directly across species.

Here we hypothesized that there could be differences in basic features of the neural code across species and brain regions, and moreover, that differences between the amygdala and the cingulate cortex would parallel evolutionary-driven differences between nonhuman primates and humans. Specifically, we looked for differences in *efficiency* and in *robustness*, as both can result from evolutionary pressure to develop cognition on one hand but preserve reliable responses to threats on the other. We used a unique opportunity to obtain single-unit recordings over long time scales (2 hours) from both structures in humans and macaques.

Previous studies have described important differences in the [ir]regularity of spike-trains, mainly in response to well-defined stimulus or action (Churchland et al., 2010; Maimon and Assad, 2009; Softky and Koch, 1993). We chose to extend on these approaches by using entropy-measures (Borst and Theunissen, 1999; Cover, 1991; Rieke et al., 1999). Entropy-rates enable a natural extension to networks of several neurons and provide a measure for the information capacity in a channel – a neuron. The long recording times allowed appropriate sampling (Treves and Panzeri, 1995) and accurate estimation (Strong, 1998). Moreover, because overall firing rates vary across regions, species and behaviors, and impact information-capacity (Rieke et al., 1999; Strong, 1998), we derived a novel measure by normalizing to the theoretical limit (Shannon, 1997), enabling direct comparable quantification independent of firing rate and instantaneous behavior. This approach has a natural interpretation as *efficiency*: in a given set of constraints due to local or global limits on activity(Baddeley et al., 1997; Niven and Laughlin, 2008), how efficient is the observed spike train compared to the maximal information-capacity.

We used several approaches to estimate robustness. We quantified the strength of all pairwise correlations and their lag as a measure for synchrony, as pairwise correlations reasonably describe local networks (Ohiorhenuan et al., 2010; Schneidman et al., 2006).

Further, we estimated the overlap in neural words, namely, how similar is the vocabulary in a pair of neurons. To do so we compared distributions of 'uttered' neural words by Jensen-Shannon-Divergence (JSD) (Lin, 2006), and further normalized it to the expected theoretical maximum (similarly to efficiency) to allow direct comparison. Overall, we compare efficiency to robustness in model-neurons and in the recorded populations and describe their tradeoff along regions and species.

Results

We analyzed single-units recorded from the amygdala and the cingulate-cortex in five Macaca fascicularis (Livneh and Paz, 2012a, b; Resnik and Paz, 2015; Taub et al., 2018a; Taub et al., 2018b) and seven human patients with pharmacologically intractable epilepsy (Gelbard-Sagiv et al., 2008; Paz et al., 2010). The dataset consisted of 747 single-neurons, 1502 pairs and 2617 triplets of simultaneously recorded neurons from the basolateral-complex (BLA) of the amygdala and Broadman areas 24,32 of the prefrontal-cortex in both species (Experimental Model and Subject Details Section). For a detailed discussion on the validity of the results in epilepsy patients see STAR Methods.

Whereas all comparisons between regions within a species were done on neurons that were simultaneously recorded in each individual, comparisons between species require careful considerations such as properties of neural activity in patients and multiple behavioral paradigms. We addressed these using several controls described below (see STAR Methods), but mainly by developing approaches that measure efficiency and robustness in a stimulus-independent manner over long-time scales, hence allowing the estimation of spike-train properties in each neuron individually independent of the context it was recorded in.

Lower efficiency in the amygdala and in non-humans

Whereas abrupt change in firing-rates is highly appropriate to estimate stimulus/task related coding and used in most neuroscience studies, we aimed to quantify and compare basic features of the information channels i.e. the neurons' spike-train, while controlling (normalizing) for firing rates and aiming to find differences that are orthogonal to it. To measure the overall capacity to transmit information in each neuron we used the entropy-rate of the complete recorded spike train. The entropy-rate increases and is bounded by the firing-rate (Fig.1A), yet firing-rate (FR) does not fully account for it (Fig.1A insets), and the firing pattern, i.e. spike times, determine the actual information capacity (Rieke et al., 1999). Therefore, to evaluate how much a neuron actually exploits its potential, we devised the *contrast-entropy* - defined as the proportion between the entropy-rate of the neuron and the maximum entropy of an analytic neuron with the same firing-rate (methods, eq' 1–3). Because firing-rate is limited in real neurons (Barlow, 2001; Niven and Laughlin, 2008), a high contrast-entropy measures how much the neuron is efficient.

We discretize spike trains into *letters* ($\Delta\tau$)ms) and define a *word* (W) by the number of letters it contains. For generality, we explore different combinations of letters ($\Delta\tau$)=1,2,4,8,16ms) and words (W=4,8,16 letters) and use each combination in all comparisons and analyses described hereafter. These 15 letter-word combinations span a wide range of words with different lengths, from 4 to 256ms, allowing us to examine the

consistency of the findings. Neurons with high contrast-entropy have word distributions that are more similar to the analytical maxima (Fig.1B top row), and those with low contrast-entropy have a substantially different distribution (Fig.1B bottom row). Due to the normalization to FR, Contrast-entropy is indeed independent of it (Fig. 1C,D, Supp.Table.1), hence allowing comparison between neurons recorded in different regions and species. Notice that as expected due to sampling data-size, as well as due to correlations between successive words (Rieke et al., 1999; Treves and Panzeri, 1995), the contrast-entropy decreases with coarser discretization (i.e. longer letters and words), and this was similar across regions and species (Supp.Fig.1).

Using the contrast-entropy to compare neurons recorded from the same region in humans and monkeys, and neurons recorded in the amygdala and the cingulate-cortex of the same species, we find that in both species cingulate neurons exhibit more efficiency than amygdala neurons (Fig.2A), and in addition, human neurons exhibit more efficiency than monkey neurons, and this was the case for both regions (Fig.2A, Wilcoxon signed-rank tests, $p<0.01$ for all, corrected for multiple comparisons). Convincingly, this was the case for the overwhelming majority of word combinations (Fig.2B; Supp.Table.1).

To establish this finding and validate it is not due to differences in firing-rates, we first notice that the order of contrast-entropy is different than the order of FR distributions across regions/species (Supp.Fig.2A,B). Further, we sampled two distributions of real neurons: in the first case, we match neurons of monkey amygdala with neurons of human amygdala with the same FR, and similarly for monkey and human cortex (Fig.3A bottom-left); and in the second, we match neurons of human amygdala with human cortex and monkey amygdala with monkey cortex (Fig.3A bottom-right). In both, the findings remain the same (Fig.3A top row, Supp.Fig.2C,D, Supp.Table.2). Finally, we adopted a ‘spike-dropping procedure’ (Fujisawa et al., 2008), and randomly removed spikes to equalize means of firing-rate distributions across regions and species (see methods), achieving similar results (Supp.Fig. 2E,F,G,H).

Although we used long periods of data (up to 2 hours) that allow reliable estimation, we validated that our findings are not affected by sampling bias (Treves and Panzeri, 1995) by further estimating the contrast-entropy for an ‘infinite word length’ (Strong, 1998) (Supp.Fig.3). The results validated again the difference between humans and monkeys for both regions, and between cingulate-cortex and amygdala in both species (Fig.3C). This was again consistent for the overwhelming majority of letter-word combinations (Fig.3C; Supp.Table.2). Finally, we made sure the results are not dependent on different recording lengths and different times during the recording, by random resampling of segments (Fig. 3D, Supp.Table.2).

We conclude that neurons have higher contrast-entropy (i.e. their spikes are more efficiently distributed) in humans compared to monkeys and in the cortex of both species compared to the amygdala.

The contribution of putative excitatory-inhibitory neurons and spike-train irregularities

A putative origin to the differences can be differential sampling and/or differential density of excitatory vs. inhibitory cell types. Although it is harder to obtain absolute cell-type identification in extracellular recordings in primates, spike waveforms can provide a reasonable proxy. The time from through to peak defined as the width of the action-potential waveforms was shown to cluster into two categories: narrow and broad, and these groups largely correspond to inhibitory interneurons and excitatory pyramidal cells, respectively (Bartho et al., 2004; Mitchell et al., 2007). We repeated this analysis for all human and monkey waveforms (all neurons that participated in this study were included), and found that in all four regions-species there are similar proportions (12–17%) of 'narrow' neurons (identified by a 'bend' algorithm applied to the cumulative-distribution-function). There was no significant difference in proportion across the four regions (χ^2, p>0.6, Supp.Fig.5B,C). Importantly, there was no correlation between the width of a neuron and its contrast-entropy, in both species (R~=0, P>0.6, Supp.Fig.5D,E). This suggests that the efficiency is not directly affected by differences in excitatory/inhibitory populations (yet it does not preclude the contribution of differential E/I balance).

Additionally, the irregularity of spike trains, traditionally measured by the coefficient of variation (CV), was shown to vary across brain regions (Maimon and Assad, 2009; Softky and Koch, 1993). Because irregularity directly contributes to the entropy-rate, we tested the contribution of the CV to our results. As expected, there is strong inverse correlation between the CV and the contrast-entropy, and as a result, we indeed find higher CV in monkeys that likely contribute to the lower contrast-entropy. However, we also observed that CV alone does not fully capture the differences we found across species and regions (Supp.Fig.6).

A tradeoff in single-neurons between efficiency and robustness

More spikes in a given time window, in a neural word in our case, increase detection of an event or stimulus and allows higher speed of response and increased reliability for a downstream region and eventually for the organism (Barlow, 2001). Therefore, increasing overall average firing rates enable both higher information transmission (Fig.1A) and at the same time higher speed and reliability of response (Fig.4). However, such increase is limited by energy consumption and bounded in neurons, a major confound for real networks (Barlow, 2001; Niven and Laughlin, 2008).

To elucidate the contribution of spike patterns, we employed model neurons generated by a two-state Markov process (Amigo et al., 2004). This model allows us to specify an entropy-rate for any particular FR by modulating only one free parameter (β) that determines the transition probability from spike to no-spike (STAR Methods). For a specific FR, a neuron can reach maximum entropy-rate when β=1, and lower values impose a reduction in the efficiency (lower entropy), but at the same time also generate more words with high spike density (this is because the maximal entropy is achieved when words with high spike density are less common, see for example Fig.1B). For each neuron in our database, we fit a β value according to its empirical FR and entropy-rate. We find higher β distribution in humans and in the cortex compared to the amygdala, in both species (Fig.4 top-right, p<0.01,

Kolmogorov-Smirnov tests). This is in agreement with the model predictions and the tradeoff hypothesis across regions and species.

In this approach, an increase in potential information comes at the expense of the speed/reliability of response, and vice versa (Fig.4). In other words, a neuron that 'wants' to maintain an overall average FR, can 'choose' to be more efficient, or instead to have higher speed/reliability of response. Indeed, when we select neurons with higher probability for words with more than one spike, we find more such neurons in the amygdala and in monkeys (Fig.4 upper-left), and when we select neurons with higher probability for words with only one spike or less, we find more such neurons in the cortex and in humans (Fig.4 lower-right).Therefore, the observed contrast-entropy reflects a tradeoff between efficiency - higher in the cortex and in humans, and robustness - higher speed/reliability of response – higher in monkeys and in the amygdala.

Higher pairwise correlations and code overlap in monkeys and in the amygdala

To test if the tradeoff between efficiency and robustness occurs also beyond single-neurons, we measured pairwise correlations. We find that neurons in the monkey and in the amygdala exhibit higher correlations compared to human and cortex, respectively (Fig.5A,B), and a higher proportion of pairs exhibited significant correlations (Supp.Table.3). To control for time-specific or task-specific contributions, we repeated the analyses with resampling segments from the recordings (Fig.5C). We further selected distributions of pairs with similar FR-differences as well as total-FR (Cohen and Kohn, 2011) and validated that the observed differences in correlations are indeed a global phenomenon (Supp.Fig.7.B,C) (Okun et al., 2015; Runyan et al., 2017).

We then quantified the lag of the cross-correlations as a measure of synchrony(Runyan et al., 2017), and find that neurons in monkeys and in the amygdala exhibit significantly shorter lags compared to humans and cortex respectively (Fig.5D,E). Such synchrony enables better downstream summation and hence a reliable population response, as well as enhanced speed-of-response. This is further in line with the hypothesis of more robust and fast/reliable response in monkeys and in the amygdala.

An additional way to measure overlaps in the code is to compare the distributions of words between pairs of neurons based on the Jensen-Shannon-Divergence (JSD) (Lin, 2006). Here as well, we find more overlaps in monkeys and in the amygdala in both species (Fig.5F, Supp.Table.4). Namely, pairs of neurons tend to use the same words more often and hence have a shared vocabulary.

A tradeoff between efficiency and robustness in non-human pairs only

The previous section described higher correlations and overlaps in pairs of neurons, interpreted as robustness of the population. In the sections before that, we characterized efficiency in single-neurons. In order to compare directly robustness with efficiency, the next step is to quantify it in pairs of neurons. To do so, we compared as before the actual entropy with the entropy of analytic pairs that have the same firing-rates. To calculate the entropy we define words and letters as before, but this time words are composed from the letters of the two neurons jointly (methods and Supp.Fig.4). Here again, the contrast-entropy of pairs was

found to be lower in monkeys than in humans, for both the amygdala and the cingulate-cortex, as well as lower in the amygdala than in the cortex of both species (Fig.6A). The same difference was revealed also when calculated for triplets of neurons recorded simultaneously (see methods and Fig.6A insets). The fact that similar findings were obtained for single neurons, pairs and triplets, strongly suggests that it is a network characteristic.

It seems reasonable to assume that in pairs, just like described for single-neurons in a previous section (Fig.4), there would be an inherent tradeoff between the contrast-entropy and the pairwise-correlations. But is such tradeoff a necessary relationship? To demonstrate that this is not the case, we shuffled neurons from different days, maintaining the contrast entropy but destroying correlations (Fig.6B right-most inset in gray). To further demonstrate this, we modeled pairs of neurons and show that we can choose β values, i.e. fixing the FR and the contrast-entropies, yet without any relationship to the cross-correlations between the surrogate neurons (Supp.Fig.7.A).

Empirically and interestingly, we find that in monkeys only there is a relationship between the contrast-entropies of pairs and their cross-correlations (Fig.6B, Pearson's correlation, $p<0.01$ for both amygdala and cortex). The difference in slopes between species was significant for both regions (Fig. 6B; $p<0.05$, Fisher Z-test). Therefore, the tradeoff between efficiency and robustness in pairs of neurons within a region is a finding unique to monkeys; or alternatively, the lack of tradeoff is unique to humans, suggesting that the local network in humans can maintain independency.

A tradeoff across species and regions

If we combine the results from the previous sections across regions and species, we observed higher efficiency in the cingulate-cortex than in the amygdala, for both primate species; and higher efficiency in humans than in non-human-primates, for both amygdala and cingulate-cortex; when efficiency is defined as efficient use of information-capacity over long spike-trains. On the other hand, we find more robustness in non-human-primates than in humans, for both regions; and in the amygdala than the cortex, in both species; when robustness is defined as the overlap in words and higher and more synchronized correlations. This is summarized in a scheme (Fig.7A).

As demonstrated in the previous section, a tradeoff between efficiency and robustness is an empirical finding and not a necessary relationship when considering pairs of neurons. To quantify if this tradeoff indeed exists across regions and species, we further plot the mean of the efficiency versus '1-robustness' for each region and species, revealing a linear relationship (Fig.7B, $p<0.01$, linear regression, error-ellipses derived from the normalized covariance matrix). The results were further validated for isolated single-units from different electrodes (Supp.Fig.7D) as well as for putative multi-unit (MUA, collapsing units across electrodes, Supp.Fig.7E), and were not a result of distance between electrodes (Supp.Fig. 7F,G,H). Finally, the tradeoff was similar when dividing recording times into neural activity surrounding presentation of external stimuli and recording times during periods without an external stimulus being presented (Fig.7C).

Discussion

We analyzed on-going neural activity from the amygdala and the cingulate-cortex of behaving humans and monkeys. We find a more efficient code (exploitation of information-capacity) in the cortex compared to the amygdala and in humans compared to monkeys. In contrast, we find that the neural code is more robust in the amygdala compared to cortex and in monkeys compared to humans. Together, it suggests an efficiency-robustness tradeoff in the neural code, and we indeed find a linear relationship between the two properties across the four examined regions. The higher efficiency in the prefrontal-cortex and in humans can potentially contribute to the higher cognitive abilities, and to the best of our knowledge, this is the first demonstration for a putative human advantage from the point of view of the neural code, in addition to the well-established neuroanatomical difference (Kaas and Herculano-Houzel, 2017). The lower robustness can also be an advantage because it allows flexibility and adaptation to changing environments, however it also comes with a cost of less reliability. Across regions, the tradeoff parallels their functional roles: the amygdala's robustness can maintain more stable emotional knowledge, namely memories that are less prone to changes or forgetting. The robustness can also contribute to faster and more reliable production of behavioral responses that are necessary for survival threats. Below we discuss the results in more detail.

We used a novel approach of the contrast-entropy as a measure. Whereas irregularities in spike-trains were compared before and were mainly used to compare stimulus-evoked responses (Churchland et al., 2010; Maimon and Assad, 2009), the contrast-entropy allowed us to characterize channels of information (neural spike trains) over long time scales, to capture higher-moments beyond the use of coefficient of variation or fano-factor, and importantly provided a direct comparable and interpretable quantity of *efficiency*. Interestingly, we found that efficiency is high overall, hovering around 90–98% in all recorded neurons in both regions and both species. These high values might point to some optimization process that already occurred along development. The consistency within a region and a species, and along the evolutionary hypothesis: from amygdala to cortex and from non-humans to humans, indicates that this might be the case. This narrow range of efficiency is directly related to the fact that most neurons use only a small range (few dozens of spikes/sec) of the theoretically possible firing rates (hundreds of spikes/sec), and together support the concept of minimizing energy consumption while gaining maximum efficiency (Baddeley et al., 1997).

On the other side of efficiency, the effect of pairwise correlations on the information in larger networks was discussed mainly in the context of stimulus-driven responses. Correlations can have a detrimental information-limiting effect (Averbeck et al., 2006), mainly due to differential-correlations (Moreno-Bote et al., 2014). Here, we aimed to capture the capacity of the neural code in on-going activity, and in this case, both efficiency of neurons and pairwise correlations constrain the effective-dimensionality (the 'intrinsic manifold') of a network (Sadtler et al., 2014) (Supp. Fig. 7I). Combining this with our findings of higher efficiency and reduced robustness, the network in humans and in the cortex potentially enables more cognitive abilities as well as flexible learning of new tasks (Golub et al., 2018). Although we show that the tradeoff between efficiency and robustness

was linear and necessary in model single-neurons (that maintain average firing-rate over long time-scales), we further show that this tradeoff is not necessary in pairs of neurons. Despite this, the empirical findings show that the tradeoff is largely linear in the recorded populations. Using pairs is a reasonable approach as pairwise correlations describe well the interactions in small networks (Ohiorhenuan et al., 2010; Schneidman et al., 2006), but likely less so in large networks with higher-order correlations (Ganmor et al., 2011; Macke et al., 2009). Therefore, we do not know how the tradeoff between efficiency and robustness behaves (analytically and empirically) for large networks. We hypothesize that at some point the tradeoff becomes necessary, unlike in pairs. Finally, we found a within-region tradeoff in monkey but not in human pairs. Although preliminary, if true, it points to another human advantage of maintaining independency of efficiency and robustness in small local networks. Obviously, increasing both efficiency and robustness is a desired feature, and even if it is limited by network size as hypothesized, the higher the independence, the better.

However, the same aforementioned considerations for improved learning in humans also suggest that ‘undesired’ learning could occur more easily (Supp. Fig. 7I). When such ‘undesired’ learning is encoded via amygdala networks due to emotional context, the higher robustness, the shared vocabulary and less efficiency can explain why such memories are less detailed (Adolphs et al., 2005), over-generalized (Dunsmoor and Paz, 2015), and harder to extinguish (Milad and Quirk, 2012). All of these cognitive characteristics can contribute to anxiety and trauma-disorders (Averbeck and Chafee, 2016; Likhtik and Paz, 2015).

In addition to the established neuroanatomical cross-species differences (Kaas and Herculano-Houzel, 2017), recent studies have shown that human neurons might have specific properties such as lower membrane capacitance (Eyal et al., 2016), enhanced dendritic compartmentalization (Beaulieu-Laroche et al., 2018) and even identified novel groups of human GABAergic interneurons (Boldog et al., 2018). Although an appealing option, we could not find any differences in the proportions of putative excitatory / inhibitory neurons between regions or species, and there was no relationship between the classification and contrast-entropy. However, this does not preclude the option that differential synaptic excitatory - inhibitory (E-I) balance contribute to the efficiency/ robustness changes. The exact contribution of input-output / dendritic-axonal organization to entropy is not well understood, and differences between primates are also unknown at this stage. Similarly, between regions, the BLA complex is a cortical-like structure cell-type wise(Carlsen and Heimer, 1988; Swanson and Petrovich, 1998), and indeed we found similar proportions of putative pyramidal-cells and interneurons. Yet one major architectural difference exists: whereas the cortex is obviously a layered structure, the BLA is likely a homogenous ‘ball’ with no clear organization (Pare and Smith, 1993; Pare et al., 1995). These structural differences most likely contribute to local correlations, shared vocabulary, and efficiency, yet how exactly they contribute and in what direction will require further modeling and in-vivo studies.

Hence, it remains unclear what is the interaction between the known neuroanatomy and the representational differences as we describe here, and whether they are dependent (Marr, 1982). From an architectural point of view, it makes sense that neurons that already sacrifice efficiency for robustness, would also cooperate among themselves to create faster and more

vigorous response in a downstream network, one that is also more resistant to noise fluctuations and hence more reliable. Indeed, cells with strong connections are those with the more correlated stimulus-related responses (Cossell et al., 2015). Robustness was also measured as overlap in neural words, i.e. the tendency of a pair of neurons to use the same short spike-patterns. The fact that there was overall less overlap in humans and cortex, is indicative and can contribute to sparse codes (Foldiak and M.P., 1998; Quiroga et al., 2005). This is further supported by the fact that it was not dependent on the distance between electrodes (unlike entropy and correlations) and the intriguing finding that there is no tradeoff in human regions only, allowing robustness to vary as required. Altogether, it suggests a more global phenomenon of using similar vocabulary to transfer information across larger modular networks.

One should carefully consider some technical aspects when interpreting the comparison between species, mainly the data from epileptic patients, different recording techniques, and different behavioral paradigms. These concerns are addressed in detail and with several controls (STAR Methods), and we provide here only the main arguments. The human data comes from epileptic patients, yet: first, only a minority (<6%) of the recordings were obtained within the epileptogenic seizure foci and ignoring these units yield the same results; second, epileptic activity is characterized by highly correlated activity in large groups of neighboring neurons, exactly the opposite of what we find in humans compared to monkeys; third, firing rates during our recording times were in the normal range, and even slightly lower in humans than in monkeys, opposite to an epilepsy concern; fourth, the patients behave normally and perform a variety of behavioral tasks during the recordings, strongly suggesting that neural coding is natural (the basic assumption behind all electrophysiological studies during behavior).

Although both human and monkey tasks required them to be attentively engaged with active responses, it might still seem difficult to interpret differences based on different behaviors across species. To address this we validated that the results hold also when repeating the analyses on randomized time periods, and separating spike-trains to periods with presentation of external stimuli from periods with no external stimulus. The validity of the results is also consistent with the finding that stimulus-induced effects are similar across regions and species, even when using different stimuli and behavioral tasks (Churchland et al., 2010). The combined datasets from several behavioral paradigms, and the novel approach we developed here designed to quantify basic properties of spike-trains over long time scales, both contribute to the argument that we unveiled task-independent general properties of the four networks. Finally, the results across amygdala vs. cingulate-cortex are a direct comparison because both were recorded simultaneously in each species in several tasks. The fact that the findings occurred independently in both species matching our original hypothesis and interpretation, provides strong support to the across species comparison.

Nevertheless, it remains as a future challenge to examine if and how the differences we describe affect instantaneous stimulus-evoked representations (Treves et al., 1999). Here, we did not aim to compare information about a specific stimulus, and it is not clear how exactly this can be done across species at all, because internal representation and behaviors (e.g.

emotional responses, associations, memories, thoughts-contexts, use of motor actuators) are much harder to control across species. Instead, we asked what general features of the neural code differentiate neural information channels in different regions and species, and focused on two complementary characteristics.

We suggest that the tradeoff we identified here across regions and species is due to evolutionary pressure that shifts the neural code from robustness, namely speed and vigor of response, both necessary for reliable execution of basic survival responses; into efficiency, enabling better exploitation of information capacity and complex use of the neural vocabulary to adapt and learn new environments. This is in line with the evolutionary transition across species, and also with the known roles of the cingulate-cortex on one hand and the amygdala on the other. We conclude that cross-species investigations are crucial for understanding basic features of the neural code, and for translational psychiatry that relies on understanding maladaptive learning and memory in neural networks.

STAR Methods

CONTACR FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rony Paz (rony.paz@weizmann.ac.il)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Non-human-primate recordings—We used data from 3 Macaca fascicularis (Livneh and Paz, 2012a, b; Resnik and Paz, 2015; Taub et al., 2018a; Taub et al., 2018b) and 2 more Macaca fascicularis Monkeys (4–7 kg) were implanted with a recording chamber (27 × 27 mm) above the amygdala under deep anesthesia and aseptic conditions. All surgical and experimental procedures were approved and conducted in accordance with the regulations of the Weizmann Institute Animal Care and Use Committee (IACUC), following NIH regulations and with AAALAC accreditation. Food, water, and enrichments (e.g., fruits and play instruments) were available *ad libitum* during the whole period, except before medical procedures that require deep anesthesia. Anatomical MRI scans were acquired before, during, and after the recording period. Anatomical images were acquired on a 3-tesla MRI scanner: (MAGNETOM Trio, Siemens) with a CP knee coil (Siemens). T1 weighted and 3D gradient-echo (MPRAGE) pulse sequence was acquired with TR of 2500 ms, TI of 1100 ms, TE of 3.36 ms, 8° flip angle, and 2 averages. Images were acquired in the sagittal plane, 192 × 192 matrix and 0.83 mm or 0.63 mm resolution. A first scan was performed before surgery and used to align and refine anatomical maps for each individual animal (relative location of the amygdala and anatomical markers such as the interaural line and the anterior commissure). We used this scan to guide the positioning of the chamber on the skull at the surgery. After surgery we performed another scan with two electrodes directed toward the amygdala, and 2–3 observers separately inspected the images and calculated the amygdala anterior–posterior and lateral-medial borders relative to each of the electrode penetrations. The depth of the amygdala was calculated from the dura surface based on the MRI at all penetration points. We used clear anatomical markers and visual similarity to identify the amygdala based on MRI images from primate atlas.

The monkeys were seated in a dark room and each day, 3–4 microelectrodes (0.6–1.2 MΩ glass/narylene-coated tungsten, Alpha Omega or We-sense) were lowered inside a metal guide (Gauge 25xxtw, OD:0.51 mm, ID:0.41 mm, Cadence) into the brain using a head-tower and electrode-positioning-system (Alpha Omega). The guide was lowered to penetrate and cross the dura and stopped ~0.5–1 cm in the cortex. The electrodes were then moved independently further into the amygdala or the cingulate-cortex (we performed 4–7 mapping sessions in each animal by moving slowly and identifying electro-physiological markers of firing properties tracking the known anatomical pathway into the amygdala). Electrode signals were preamplified, 0.3 Hz-6 KHz bandpass filtered and sampled at 25Khz; and on-line spike sorting was performed using a template-based algorithm (Alpha Lab Pro, Alpha Omega). We allowed 30 min for the tissue and signal to stabilize before starting acquisition and behavioral protocol. At the end of the recording period, off-line spike sorting was further performed for all sessions to improve unit isolation (offline sorter, Plexon Inc).

Human recordings—We used data from 7 patients with pharmacologically intractable epilepsy that have well isolated neurons from the amygdala or the anterior cingulate cortex (Gelbard-Sagiv et al., 2008; Paz et al., 2010). Extensive noninvasive monitoring did not yield concordant data corresponding to a single respectable epileptogenic focus. Therefore, they were implanted with chronic depth electrodes for 7–10 days to determine the seizure focus for possible surgical resection. All studies conformed to the guidelines of the Medical Institutional Review Board at University of California at Los Angeles. The electrode locations were based exclusively on clinical criteria and were verified by MRI or by computer tomography coregistered to preoperative MRI. Each electrode consisted of a flexible polyurethane probe containing nine 40-μm platinum–iridium microwires protruding ~4 mm into the tissue beyond the tip of the probe. Eight microwires were active recording channels and referenced to the ninth, lower impedance, microwire. The differential signal from the microwires was amplified by using a 64-channel Neuralynx™ system, filtered between 1 and 9,000 Hz and sampled at 28 kHz. All sessions were conducted at the patients' quiet bed-side using a standard laptop screen and speakers. Spike detection and sorting was applied to the continuous recordings by using a well-established clustering algorithm (Quiroga et al., 2004). After sorting, the clusters were classified into single units or multi units based on: (i) the spike shape and its variance; (ii) the ratio between the spike peak value and the noise level; (iii) the ISI distribution of each cluster; and (iv) the presence of a refractory period for the single units. Only well isolated neurons were considered to the further of the analysis

Sample size - numbers of single-units, pairs and triplets of neurons.

	Human Amygdala	Human Cortex	Monkey Amygdala	Monkey Cortex	Total
Single	80	42	260	365	747
Pairs	313	152	369	668	1502
Triplets	898	378	479	862	2617

Demographic information – age and sex of all patients and monkeys

Identifier	Age	Sex	Species
P384	25	F	Human
P385	21	F	Human
P386	18	F	Human
P394	30	M	Human
P395	44	M	Human
P399	N/A	F	Human
P400	46	M	Human
‘L’	5	M	Monkey
‘B’	6	M	Monkey
‘Z’	6	M	Monkey
‘D’	5	M	Monkey
‘G’	6	M	Monkey

We were not able to compare our measures separately between human males and females due to marginal statistical power.

METHOD DETAILS

Behavioral paradigms—Three of the five monkeys underwent discriminatory classical conditioning, reversal, and extinction, using different neutral tones or pictures every day that acquired value through conditioning with either appetitive or aversive odors(Livneh and Paz, 2012a, b; Taub et al., 2018a; Taub et al., 2018b). The monkeys were seated in a chair with a custom-made nasal mask attached to their nose. The mask was attached to two pressure sensors with different sensitivity range that enable real-time detection of breath onset. Experimental sessions initiated by a habituation session of ten presentations of the CS. The acquisition session that followed included 30 trials of CS paired with an aversive odor. Propionic acid stimulates olfactory and trigeminal receptors at the nose and is highly aversive to humans and monkeys. CS was triggered by breath onsets, and odor (US) was released at the following breath onset (but not before 1 s elapsed). Twenty unpaired CSs were presented to the monkey in order to extinguish the acquired association between the CS and the US.

Two of the monkeys underwent a perceptual 2-alternative-forced-choice task, where visual stimulus was presented for either 30, 60, 130, 230 or 330 ms in different trials to vary difficulty. Sets of stimuli changed on a daily basis to induce daily learning. Responses were delivered by pressing left/right buttons, and correct responses entailed liquid reward. The monkeys were seated in a chair in front of a monitor with a three-buttons panel located below it. Each trial was initiated by the monkey holding the middle button, followed by a delayed quick presentation of a visual stimulus that indicated the response they should perform to receive a rewarding water drop, pressing either left or right button.

The seven humans engaged in two alternating tasks, one included viewing several short clips, and the other involved free recall of the viewed content (without external stimulus)

(Gelbard-Sagiv et al., 2008; Paz et al., 2010). Each recording session was composed of 1–3 (average 1.6) iterations of two parts: In the viewing session, subjects were presented with a series of between 10 to 16 different audiovisual movie-clips lasting 5 to 10 seconds each. Each clip depicted an "episode" featuring famous people or characters engaged in activity, landmarks photographed from various views, animals in motion, or objects depicted in a dynamic context. Each clip was presented 5–10 times and order of presentation was pseudorandomized: each cycle contained all different clips, but order of clips was randomized within the cycle; same clip was never presented twice consecutively; all clips within a single session were of same length; in some of the experiments interleaving blank periods ("blanks") of 5 s were used occasionally within a group of successive clips, and in other experiments interleaving blanks of 2–3 s were used before each clip. Patients were asked to freely watch the clips. In the free recall session that followed, patients were asked to freely recall the clips they had just seen and verbally report immediately when a clip "comes to mind". This session was not limited in time and was stopped only when the patient recalled correctly all the clips or when the patient could not remember any more clips.

All tasks kept the participants (monkeys and humans) highly engaged, and involved external stimuli as well as required active voluntary responses. Please see text and methods for the rationale to use a diverse set of behaviors to increase robustness and validity of findings.

Contrast-entropy—The total entropy of the spike train, which quantifies the variations across time and sets the capacity of the spike train to carry information, was estimated in the following way (Strong, 1998),(Rieke et al., 1999) : the neural spike train is discretized into bins of size $\Delta\ \tau$ ms, and we refer to those bins as binary *letters*. A letter is equal to one if during the period of $\Delta\ \tau$ at least one spike occurred, or zero otherwise. We define a *word* (W) in the neural code by the number of letters that it contains. Therefore, the length of the word is: $T = W\ \Delta\ \tau$. The number of each word occurrences is then normalized by the total number of words, to get its occurrence probability p_i. The entropy rate of a neuron is calculated by:

$$Entropy_D = \frac{-\sum_i p_i log_2 p_i}{T} \quad (1)$$

In equation (1), we divided the total entropy by T in order to get the entropy rate. In order to compare neurons across regions and species, and due to the high-dependence on firing rate (See text and Fig.1), we normalized each entropy-rate to the maximum entropy that can be obtained from of a spike train with the same firing rate. Maximum entropy is obtained when we consider every spike as a random event. Therefore, the probability of a spike was calculated according to the mean firing rate of the neuron ($\bar{r}$), $p_S = \bar{r}\ \Delta\ \tau$. Then, the analytic entropy rate equals to:

$$Entropy_A = \frac{-\sum_{i=0}^{W} C_i^W p_S{}^i (1-p_S)^{W-i} log_2\left(p_S{}^i (1-p_S)^{W-i}\right)}{T} \quad (2)$$

Where $C_i^W = \frac{W!}{(W-i)!i!}$.. This analytic expression yields the same result as in (Rieke et al., 1999).

To get an estimation of the entropy rate that is independent on the firing rate, and measures how much a neuron exploits its potential to transfer information, we define the contrast-entropy as follows:

$$ContrastEntropy = \frac{Entropy_D}{Entropy_A} \quad (3)$$

Information measures, and entropy-rate included, are prone to bias due to limited data sampling (Treves and Panzeri, 1995). To correct for this, we further estimated the contrast-entropy based on the approach of (Strong, 1998). To do so, the naïve entropy is plotted for different Words, while $\Delta\tau$ is held constant. Extrapolation for infinite word length is achieved when the regression line fitted to all words intercepts the y-axis (Supp.Fig.3). The estimated entropy rate for each neuron was then used (instead of the naïve entropy rate as in eq' 3) to calculate the estimated contrast-entropy for the different letter lengths.

Two-state Markov process neurons—To better understand how neurons distribute their spikes to achieve a specific entropy rate, yet under physiological constraints of overall average firing rates, we modelled and simulated neurons that can have a specific firing rate and specific entropy rate. To do so, we used a two-state Markov process, where two transition probabilities are defined (Amigo et al., 2004; Cover and Thomas, 2006).

P_{10} – the probability to shift from no-spike to spike; and P_{01} – the probability to shift from spike to no-spike. The probabilities $(P_{11},\ P_{00})$ are derived, by definition, by summing the complementary probabilities to one.

Therefore, the stationary solution is:

$$P_0 = \frac{P_{01}}{P_{01}+P_{10}};\ P_1 = \frac{P_{10}}{P_{01}+P_{10}} \quad (4)$$

For a specific firing rate (FR), the proportion of the probabilities of the stationary solution should be constant and equal to:

$$\frac{FR}{1000} = \frac{P_1}{P_0 + P_1} = \frac{P_{10}}{P_{01} + P_{10}} \quad (5)$$

This leaves one degree of freedom that we could use to get the required entropy rate of the neuron. The maximum entropy is achieved when $P_{01} = 1 - \frac{F_R}{1000}$ (see proof in this methods under "Maximum entropy of two-state Markov process – Proof"). The entropy rate of such a process generating an infinite long binary sequence is given by (Cover and Thomas, 2006), (Amigo et al., 2004):

$$H_m = P_0\{-P_{10}log_2(P_{10}) - (1-P_{10})log_2(1-P_{10})\} \quad (6)$$
$$+ P_1\{-P_{01}log_2(P_{01}) - (1-P_{01})log_2(1-P_{01})\}$$

Therefore, we choose $P_{01} = \left(1 - \frac{F_R}{1000}\right)\beta$, where (a coefficient range between 0 to 1) allows to change and choose the entropy rate of the neuron, while maintaining the firing rate constant.

Then, for each real neuron, we fit the β that minimizes the difference between the entropy rate of the neuron and the surrogate neuron H_m with the same firing rate.

Overlaps in words by Jensen-Shannon Divergence (JSD)

The Jensen–Shannon divergence quantifies the dissimilarity of the distributions p and q:

$$JSD = \frac{1}{2}D_{KL}\left(p, \frac{p+q}{2}\right) + \frac{1}{2}D_{KL}\left(q, \frac{p+q}{2}\right) \quad (7)$$

Where D_{KL} is the Kullback-Leibler divergence defined as $D_{KL}(p,q) = \sum_x p(x)log_2\frac{p(x)}{q(x)}$

Calculation of JSD from the data is done by estimating the probability distribution of words for each neuron in a simultaneously recorded pair of neurons (as done for the entropy).

Using the probability of spike of the first neuron p_S (based on its FR), the probability of spike of the second neuron q_s and the word's length W, we can derive the analytic JSD:

$$JSD_A = \sum_{i=0}^{W}[\frac{C_i^W}{2}(p_S{}^i(1-p_S)^{W-i}log_2\left(\frac{2p_S{}^i(1-p_S)^{W-i}}{p_S{}^i(1-p_S)^{W-i} + q_S{}^i(1-q_S)^{W-i}}\right) \quad (8)$$
$$+ q_S{}^i(1-q_S)^{W-i}log_2\left(\frac{2q_S{}^i(1-q_S)^{W-i}}{p_S{}^i(1-p_S)^{W-i} + q_S{}^i(1-q_S)^{W-i}}\right)]$$

And the Contrast JSD (in analogy to contrast-entropy), is therefore:

$$ContrastJSD = \frac{JSD_D}{JSD_A} \quad (9)$$

The contrast JSD is highly dependent on the difference between the firing rates of the cells. In order to compare contrast JSD, we choose pairs of neurons with similar difference in firing rate (separately for each region, Supp.Fig.7B,C). The contrastJSD is 1 if the JSD of the data is exactly like the analytical JSD, and for presentation purposes (in Fig.5F only), we normalize and define contrastJSD = Abs(contrastJSD-1).

Pairwise correlations—Pairwise correlations were calculated in the traditional way by using Pearson correlations for the discretized pairs of neurons, for every word and letter combination.

The time-lag of maximal correlation is calculated by time shifting one neuron in respect to the first neuron, and finding the optimal lag for each pair. This is equivalent to identifying the peak location in a classical cross-correlation (see Fig.5)

Entropy for pairs of neurons—The entropy of pairs of neurons is calculated by discretizing each spike train into bins of size $\Delta\ \tau$, and taking word W from each neuron to create a joint word in the length $2 * W$ (see Supp.Fig.4) The analytic probability of a word with i spikes from the first neuron, and j spikes from the second is:

$$P_{pair}(i,j) = {p_{S1}}^{i}(1-p_{S1})^{W-i}{p_{S2}}^{j}(1-p_{S2})^{W-j} \quad (10)$$

Therefore, the analytic entropy of pairs equals to:

$$ENT_{pair} = -\sum_i \sum_j C_i^W C_j^W P_{pair}(i,j) log_2 P_{pair}(i,j) \quad (11)$$

The entropy rate of pairs from the data is calculated by estimating the words' distribution in the same way it was done for single neurons, but by combining the i^{th} word of the first neuron to the j^{th} word of the second neuron.

The naïve contrast-entropy in pairs is therefore:

$$ContrastEntropy2 = \frac{Entropy_D pair}{ENT_{pair}} \quad (12)$$

As for single-neurons, we further validated the estimated contrast-entropy2 for pairs by using the method of (Strong, 1998).

Entropy for triplets of neurons—The entropy of triplets of neurons is calculated by discretizing each spike train into bins of size $\Delta\ \tau$, taking word W from each neuron to create a joint word in a length of $3 * W$.

The analytic probability of a word with i spikes from the first neuron, j spikes from the second and k spikes from the third neuron:

$$P_{triplets}(i,j,k) = {p_{S1}}^{i}\left(1-p_{S1}\right)^{W-i}{p_{S2}}^{j}\left(1-p_{S2}\right)^{W-j}{p_{S3}}^{k}\left(1-p_{S3}\right)^{W-k} \quad (13)$$

Therefore, the entropy of triplets equals to:

$$ENT_{triplets} = -\sum_i \sum_j \sum_k C_i^W C_j^W C_k^W P_{triplets}(i,j,k) log_2 P_{triplets}(i,j,k) \quad (14)$$

And similar to singles and pairs, the contrast entropy of triplets is:

$$ContrastEntropy3 = \frac{Entropy_D triplets}{ENT_{triplets}} \quad (15)$$

Choosing neurons and pairs with similar firing rate—To further control and compare across species and regions independent of their firing rate, we created samples of neurons with similar properties. If we have two groups of neurons, the first one contains N1 neurons, and the second contains N2 neurons, we examine which group is smaller:

$$Ns = N1\ if\ N1 \leq N2 \ \text{ or } \ Ns = N2\ if\ N1 > N2$$

We rank Ns neurons in the largest group according to the smallest difference from the closest neuron in the smallest group (FR-wise). If the difference is larger than 0.1Hz, the neurons are ignored in both groups. The result is two groups with the same number of neurons and with differences in firing rate smaller than 0.1Hz.

This process was done across regions (human amygdala with monkey amygdala and human cortex with monkey cortex) and across species (human amygdala with human cortex and monkey amygdala with monkey cortex).

We apply the same method to compare differences between pairs of neurons in each group (across species and across regions) to create distributions of pairs with smallest differences in firing rates and in the sums of firing rates (Supp.Fig.7B,C).

Correlation between contrast entropy of single neuron and pairwise correlations—Every simultaneously recorded pair of neurons was used. For every neuron we calculate the contrast-entropy as in eq' 3 and the sum of contrast-entropies:

$$SumContrast = ContrastEntropy(1) + ContrastEntropy(2) = \frac{Entropy_{D1}}{Entropy_{A1}} + \frac{Entropy_{D2}}{Entropy_{A2}} \quad (16)$$

The relationship between the SumContrast and the cross-correlation was estimated by Pearson-coefficient.

Fisher r-z transformation was used to test for significant differences across regions and species:

$$T_1 = \frac{log\left(\frac{1+R_1}{1-R_1}\right)}{2},\ T_2 = \frac{log\left(\frac{1+R_2}{1-R_2}\right)}{2} \quad (17)$$

$$z = \frac{T_1 - T_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}} \quad (18)$$

Where N_1 and N_2 are the number of pairs that were used in each species and region, followed by a z-test for significance.

Shuffling pairs—To test if the relationship between the sum of contrast entropies and the cross-correlation is a necessary one (i.e. pairs with low sum of contrast-entropy are expected to have high cross-correlations), we shuffled neurons to use pairs recorded in different days.

External stimuli vs. no-stimuli recording periods—We divided the data into two separate parts of each paradigm to represent two different states in each species: presentation of external stimuli and no-stimulus being presented. For humans, the two states are (Gelbard-Sagiv et al., 2008; Paz et al., 2010): 1. During clip viewing; and 2. During free recall without any external stimulus. In monkeys, the two states are(Livneh and Paz, 2012a): 1. During the CS-US presentation of a trace-conditioning task (CS is a tone and US is an aversive odor or liquid reward); 2. During long inter-trial-interval periods without external stimulus. Although it Is admittedly hard to know what internal process is ongoing in each species in each phase, the high similarity of the results across states and independency of the main finding (tradeoff between robustness and efficiency) from the task/state, strongly suggest that the results are general and do not depend on differences in behavioral paradigms or internal states (see bellow).

Ellipses and Intrinsic manifold—In order to estimate the overall differences between the tradeoffs across species and regions, we calculated the mean and the error ellipse for

each region. The lengths of the ellipse axes are the square root of the eigenvalues of the normalized covariance matrix. The center of the ellipses is the mean of the X (efficiency) and Y (1-robustness) values.

The projection of each mean (as a vector from the origin to the center of each ellipse) onto the identity line is proportional to the dimension of the 'intrinsic manifold'. This is ofcourse constrained by our ability to measure only 2^{nd}-order correlations in the current data.

Coefficient of Variation (CV)—Coefficient of variation (CV) usually provide a parameter-free method to describe the inter spike interval (ISI) distribution. CV, a measure of the irregularity of the spike train, is defined as the standard deviation of the ISI divided by its mean. We first calculated the CV of all the spike train, but because we had long recording times and the firing rate of cells can vary over time, the ISI histogram can resemble the sum of more basic distributions. To address this, we adopted a method for plotting six ISI histograms per neuron in which each histogram consists of intervals associated with a similar mean firing rate(Maimon and Assad, 2009; Softky and Koch, 1993). We divided each spike train into certain time windows, and calculated the local firing rate in each window.

T_s - the time length of the spike train; W_i – is the window, therefore there are $N = \frac{T_s}{W_i}$ windows in the spike train. We divided those N windows into 6 groups by their local firing rates, such that first group had $\frac{N}{6}$ windows with the lowest firing rates, and so on. We plotted the ISI distribution for each group and calculated its CV. The CV of the spike train is therefore the mean CV of those 6 groups. See Supp.Fig.6.

Random spike-dropping-procedure—Inspired by the "thinning" procedure in (Fujisawa et al., 2008), we randomly removed spikes from the spike trains in order to compare the mean of the firing rate distributions across regions (Amygdala to CC in both species) and across species (Humans to monkeys in both regions). The percentage of spikes that were removed defined by the differences in the mean of the firing rate distributions across the groups (see Supp.Fig.2; Spikes were mostly removed from monkey amygdala neurons when compared with human amygdala or monkey CC and from human CC when compared to human amygdala). The percentage of spikes removed was equal for all neurons. Altogether it allowed us to create groups with equal mean firing rate while maintaining the distribution of spikes as similar to the origin as possible.

Considerations regarding the comparison across species

A. The neurological pathology of the patients.: Several important considerations that convince us it does not affect the main results:

- Only a minority of the recordings (less than 6% of the units) were obtained from within the epileptogenic seizure foci. Moreover, when we repeated the analyses ignoring these units, the conclusions did not change.

- Epileptic activity is characterized by highly correlated activity in large groups of neighboring neurons. This is exactly the opposite of what we find (higher correlations in monkeys than in humans).
- Firing rates during our recording times were in the normal range (Supp.Fig.2). They were even slightly lower in humans than in monkeys, opposite to an epilepsy concern (which exhibit in high rates), and it is also opposite to our main findings (higher entropy in humans).
- Previous studies have demonstrated high correlation between single unit activity in epileptic patients and fMRI BOLD signal in normal subjects (Mukamel et al., 2005), suggesting that the neuronal activity of epileptic patients - in the absence of seizures - is not fundamentally different from normal.
- The patients are completely drug free during the week or two of the recordings (for clinical reasons, to enable observation of natural neural activity).
- The neuroscience community widely accepted many findings and knowledge about coding in the human brain based on recordings in such patients, published extensively in high-profile journals (e.g. Quiroga et al. Nature 2005; Gelbard-Sagiv et al. Science 2008; Paz et al, PNAS 2010; Tang et al. Neuron 2014; Kaminsi et al. Nature neuroscience 2017; and many more).
- The patients behave normally and perform a variety of normal behavioral tasks during the recordings, strongly suggesting that neural coding during recordings is natural (notice this is the basic assumption behind all neuroscience studies in any animal, without it no interpretation can be made on electrophysiological studies).
- Finally, the main finding was found across regions in both species separately, matching our hypothesis and final interpretation. Of course, recordings were simultaneous and with identical techniques within a species, and all procedures performed by the same person within a species (i.e. for both regions). Although this is not a complete proof for the cross-species finding, the fact that it was found twice in an independent manner, provides additional support in our view to the cross-species finding as well.

B. Different recording systems across species: Units in humans were recorded using macro-electrode that contain nine 40-μm platinum–iridium micro-wires protruding ~4 mm into the tissue beyond the tip of the probe. In monkeys, units were recorded using single micro-electrodes of glass/narylene-coated tungsten.

We performed several analyses to show it is highly unlikely to account for the main findings. We first refer to the correlations which are most relevant for this concern in our view:

- First, we validated in our data that correlations indeed decrease with distance between electrodes, as expected from classical findings (Supp. Fig. 7).
- However, the distance between electrodes in monkeys is actually larger than in humans, due to the average distance between contacts on the depth electrode used in humans vs. the average distance between electrodes inserted at different

locations in the grid we use (above the skull) in monkeys. So in fact, this should work 'against' us. Despite this, correlations are higher in monkeys (the main finding of robustness).

- To further validate this is the case, we re-performed the analyses taking only one neuron per-electrode in monkeys and in humans, which enforces a larger distance on all pairs of simultaneously recorded neurons in monkeys. We found that the correlations remain higher than in humans (Supp.Fig.7)
- Similarly, we find that the contrast-entropy (for pairs) increases with distance between neurons, and this is again opposite to our main finding that it is lower in monkeys (as aforementioned, distance between electrodes is larger in monkeys).
- The JSD show no relationship to distance between electrodes (Supp.Fig.7), suggesting the main finding does not stem from this factor.

We show full distributions of firing rates for all four regions (Supp.Fig.2). They were all within a reasonable range, and differences within species (across regions) were very similar to differences across species. Moreover, the differences in FR across regions and species do not show the same trend shown by our main results (for contrast-entropy, robustness, and tradeoff). Finally, our analyses were designed specifically to account for the FR and this is supplemented by several other methods which validate our findings. To conclude, firing-rates cannot account for our findings.

The last concern is that of unit-isolation. We provide arguments why it is highly unlikely it accounts for the findings:

- First and importantly, changes in unit-isolation (mainly due to recording techniques but that also affect spike-sorting) mainly affect spiking characteristics. These influence mainly the FR, and hence are not an issue here (as described in the previous paragraph). Given that the FR is not a concern here, below we describe further controls for potential differential unit-isolation:
- We repeated the analyses but this time taking only one unit per electrode, hence making sure there are no overlaps (this is especially important for the correlations, as mentioned above, but also for the other analyses. Supp.Fig.7).
- The complement is the case where unit-isolation is less precise in one place vs. another, and hence more single-units are actually MUA. This is a bigger concern in the human data than in the monkey data, and we therefore combine/collapse single-units across electrodes in monkey data (to create on purpose MUA), and re-analyze. We find similar results (Supp.Fig.7).
- The distributions of spike waveforms (when considering the main criteria for excitatory-inhibitory separation) yield similar results across regions and species, both in shape of the distributions (CDF) and in the cutoff for E/I. This strongly suggests against a large bias in unit-isolation due to waveforms (Supp.Fig.5).
- We emphasize that the tradeoff was found across regions in both species separately, matching our hypothesis and final interpretation. Of course,

recordings were simultaneous and with identical techniques within a species, and all procedures performed by the same person within a species (i.e. for both regions). Although this is not a complete proof for the cross-species finding, the fact that it was found twice in an independent manner, provides additional support in our view to the cross-species finding as well.

C. Different behavioral paradigms across species: We actually think it can be one of the strengths of our study, because the analyses were performed on a variety of tasks in both species and with methods designed precisely to address the generality of the efficiency/robustness measures. In addition, we took several approaches to validate that the different behavioral paradigms are not the source of the differences:

- The results remain valid when we use resampling for time segments that are randomly selected from the recording periods (Supp.Table.2 and main Fig.3D and Fig5.C).
- The tradeoff was similar when dividing recording times into neural activity surrounding presentation of external stimuli and recording times during periods without an external stimulus being presented (Fig.7C). The findings (tradeoff) remain similar, and changes due to the stimulus are far smaller than changes across species and across regions.
- Let us assume that we would do an experiment with similar behavior across humans and monkeys, can we assume that humans have exactly the same internal responses (context-based, cognitive-based, memory-based) following a stimulus as monkeys exposed to the same stimulus? In other words, this concern is inevitable when comparing different species, and the only way to address it is to plan the analyses properly to unveil differences that cannot be explained by it i.e. that are independent of the specific task. This is exactly what we did here, and this is why we developed new measures that do not rely on stimulus-evoked responses but measure efficiency and robustness over long time-scales.
- We emphasize again that the tradeoff was found across regions in both species separately, matching our hypothesis and final interpretation. The fact that it was found twice in an independent manner, and according to our hypothesis and interpretation, provides additional support in our view to the cross-species finding as well.
- In a study that compared stimulus vs. no-stimulus (Churchland et al., 2010), they observed similar properties across region/species. We show here that the efficiency-robustness tradeoff across species/regions - is similar for stimulus and non-stimulus. They also used different stimuli across species (and even regions) to show similar effects.

We cannot, in the current case, and did not aim to, compute how much information a neuron holds per stimulus. This can be an interesting question in itself to compare across species, yet if the finding was stimulus-specific, it would have been a different question altogether (as it could be a result of many factors: context, memory, strategy and so forth). Overall, we

see our approach as a necessary complementary one (in addition to the traditional ‘stimulus driven’). The main finding is about the tradeoff between efficiency and robustness in long spike-trains. We looked for basic properties that are different across species and regions in a general manner.

Maximum entropy of two-state Markov process - Proof—The entropy to maximize is:

$$H_m = P_0\{-P_{10}log_2(P_{10}) - (1-P_{10})log_2(1-P_{10})\} + P_1\{-P_{01}log_2(P_{01}) - (1-P_{01})log_2(1-P_{01})\}$$

Express all the variables as function of F_R and P_{01}:

$$P_{10} = \frac{P_{01}F_R}{1000 - F_R}$$

$$P_0 = \frac{P_{01}}{P_{01}+P_{10}} = \frac{P_{01}}{P_{01} + \frac{P_{01}F_R}{1000-F_R}} = \frac{1000-F_R}{1000}$$

$$P_1 = \frac{P_{10}}{P_{01}+P_{10}} = \frac{\frac{P_{01}F_R}{1000-F_R}}{P_{01} + \frac{P_{01}F_R}{1000-F_R}} = \frac{F_R}{1000}$$

$$H_m = \frac{1000-F_R}{1000}\left\{-\frac{P_{01}F_R}{1000-F_R}log_2\left(\frac{P_{01}F_R}{1000-F_R}\right) - \left(1-\frac{P_{01}F_R}{1000-F_R}\right)log_2\left(1-\frac{P_{01}F_R}{1000-F_R}\right)\right\} + \frac{F_R}{1000}\{-P_{01}log_2(P_{01}) - (1-P_{01})log_2(1-P_{01})\}$$

For finding the value, let us find when the derivative equal to zero:

$$\frac{\partial H_m}{\partial P_{01}} = \frac{F_R}{1000}\left\{-log_2(P_{01}) + log_2(1-P_{01}) - log_2\left(\frac{P_{01}F_R}{1000-F_R}\right) + log_2\left(1-\frac{P_{01}F_R}{1000-F_R}\right)\right\} = 0$$

And the last equation is correct when:

$$P_{01} = 1 - \frac{P_{01}F_R}{1000-F_R}$$

Therefore, the entropy is maximized for:

$$P_{01} = 1 - \frac{F_R}{1000}$$

QUANTIFICATION AND STATISTICAL ANALYSIS—The number of neurons (n) that were used in this study across regions and species is described in Experimental model and subject details section.

Several custom-written MATLAB codes were used:

Contrast-entropy calculations of single neurons were based on equations (1)–(3) in the methods. Two-state-Markov neurons were simulated based on equations (4)–(6) in the methods. Contrast JSD was calculated based on equations (7)–(9). Contrast-entropy of pairs and triplets were calculated based on equations (10)–(15). A code for choosing neurons with similar FR was written as described in methods. A code for drawing the ellipses and intrinsic manifolds was written as described in the methods. CV calculations are based on the derivations in (Maimon and Assad, 2009), and described also in the methods. Random spike-dropping procedure is based on (Fujisawa et al., 2008) and described in the methods.

All statistical tests were conducted in Matlab. The statistical tests are described in the main text, methods and relevant figure legend, specifically:

The correlation coefficient and P-values in Fig.1D are calculated based on Pearson correlation coefficient, and differences between entropy-rate and contrast entropy in Fig.1E are tested with Fisher z-test. Differences across regions and species (Fig.2, Fig.3., Fig.6 and their Supp. corresponds) were tested using Wilcoxon signed-tank test. Differences in cumulative-density-functions (CDFs) across species and regions (Fig.4, Fig.5) are tested using Kolmogorov-Smirnov test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Yossi Shohat for supervising animal welfare and experimental procedures; Drs. Uri Livneh, Jennifer Resnik and Aryeh Taub for providing experimental data, Dr. Eilat Kahana for help with medical and surgical procedures; Dr. Edna Furman-Haran, Nachum Stern and Fanny Attar for MRI procedures. This work was supported by Grants from the National Institute of Neurological Disorders and Stroke (NINDS no. R01NS033221 and R01NS084017) to I. Fried; a Schaefer scholar award (Columbia Univ.), ISF #26613 and ERC-2016-CoG #724910 grants to R. Paz.

References

Adolphs R (2010). What does the amygdala contribute to social cognition? Ann N Y Acad Sci 1191, 42–61. [PubMed: 20392275]

Adolphs R, Tranel D, and Buchanan TW (2005). Amygdala damage impairs emotional memory for gist but not details of complex stimuli. Nat Neurosci 8, 512–518. [PubMed: 15735643]

Amigo JM, Szczepanski J, Wajnryb E, and Sanchez-Vives MV (2004). Estimating the entropy rate of spike trains via Lempel-Ziv complexity. Neural Comput 16, 717–736. [PubMed: 15025827]

Averbeck BB, and Chafee MV (2016). Using model systems to understand errant plasticity mechanisms in psychiatric disorders. Nat Neurosci 19, 1418–1425. [PubMed: 27786180]

Averbeck BB, Latham PE, and Pouget A (2006). Neural correlations, population coding and computation. Nat Rev Neurosci 7, 358–366. [PubMed: 16760916]

Baddeley R, Abbott LF, Booth MC, Sengpiel F, Freeman T, Wakeman EA, and Rolls ET (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. Proc Biol Sci 264, 1775–1783. [PubMed: 9447735]

Barlow H (2001). Redundancy reduction revisited. Network 12, 241–253. [PubMed: 11563528]

Barrickman NL, Bastian ML, Isler K, and van Schaik CP (2008). Life history costs and benefits of encephalization: a comparative test using data from long-term studies of primates in the wild. J Hum Evol 54, 568–590. [PubMed: 18068214]

Bartho P, Hirase H, Monconduit L, Zugaro M, Harris KD, and Buzsaki G (2004). Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. Journal of neurophysiology 92, 600–608. [PubMed: 15056678]

Barton RA, and Aggleton JP (2000). Primate evolution and the amygdala In The amygdala: a functional analysis, Aggleton JP, ed. (Oxford U. Press).

Beaulieu-Laroche L, Toloza EHS, van der Goes MS, Lafourcade M, Barnagian D, Williams ZM, Eskandar EN, Frosch MP, Cash SS, and Harnett MT (2018). Enhanced Dendritic Compartmentalization in Human Cortical Neurons. Cell 175, 643–651.e614. [PubMed: 30340039]

Boldog E, Bakken TE, Hodge RD, Novotny M, Aevermann BD, Baka J, Bordé S, Close JL, Diez-Fuertes F, Ding S-L, et al. (2018). Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. Nature neuroscience 21, 1185–1195. [PubMed: 30150662]

Borst A, and Theunissen FE (1999). Information theory and neural coding. Nature neuroscience 2, 947–957. [PubMed: 10526332]

Byrne RW, and Corp N (2004). Neocortex size predicts deception rate in primates. Proc Biol Sci 271, 1693–1699. [PubMed: 15306289]

Carlsen J, and Heimer L (1988). The basolateral amygdaloid complex as a cortical-like structure. Brain Res 441, 377–380. [PubMed: 2451985]

Chater N, Tenenbaum JB, and Yuille A (2006). Probabilistic models of cognition: conceptual foundations. Trends Cogn Sci 10, 287–291. [PubMed: 16807064]

Churchland MM, Yu BM, Cunningham JP, Sugrue LP, Cohen MR, Corrado GS, Newsome WT, Clark AM, Hosseini P, Scott BB, et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nature neuroscience 13, 369–378. [PubMed: 20173745]

Cohen MR, and Kohn A (2011). Measuring and interpreting neuronal correlations. Nat Neurosci 14, 811–819. [PubMed: 21709677]

Cossell L, Iacaruso MF, Muir DR, Houlton R, Sader EN, Ko H, Hofer SB, and Mrsic-Flogel TD (2015). Functional organization of excitatory synaptic strength in primary visual cortex. Nature 518, 399–403. [PubMed: 25652823]

Cover TM, and Thomas JA (2006). Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing) (Wiley-Interscience).

Cover TT, sample JA (1991). Elements of Information Theory (Wiley-Interscience).

Dunsmoor JE, and Paz R (2015). Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. Biol Psychiatry.

Duvarci S, and Pare D (2014). Amygdala microcircuits controlling learned fear. Neuron 82, 966–980. [PubMed: 24908482]

Etkin A, Buchel C, and Gross JJ (2016). Emotion regulation involves both model-based and model-free processes. Nat Rev Neurosci 17, 532.

Eyal G, Verhoog MB, Testa-Silva G, Deitcher Y, Lodder JC, Benavides-Piccione R, Morales J, DeFelipe J, de Kock CP, Mansvelder HD, et al. (2016). Unique membrane properties and enhanced signal processing in human neocortical neurons. Elife 5.

Foldiak P, and M.P. Y (1998). Sparse coding in the primate cortex In The handbook of brain theory and neural networks, Michael AA, ed. (MIT Press), pp. 895–898.

Fujisawa S, Amarasingham A, Harrison MT, and Buzsáki G (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. Nature neuroscience 11, 823. [PubMed: 18516033]

Gabi M, Neves K, Masseron C, Ribeiro PF, Ventura-Antunes L, Torres L, Mota B, Kaas JH, and Herculano-Houzel S (2016). No relative expansion of the number of prefrontal neurons in primate and human evolution. Proc Natl Acad Sci U S A 113, 9617–9622. [PubMed: 27503881]

Ganmor E, Segev R, and Schneidman E (2011). Sparse low-order interaction network underlies a highly correlated and learnable neural population code. Proc Natl Acad Sci U S A 108, 9679–9684. [PubMed: 21602497]

Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, and Fried I (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. Science 322, 96–101. [PubMed: 18772395]

Ghashghaei HT, Hilgetag CC, and Barbas H (2007). Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. Neuroimage 34, 905–923. [PubMed: 17126037]

Golub MD, Sadtler PT, Oby ER, Quick KM, Ryu SI, Tyler-Kabara EC, Batista AP, Chase SM, and Yu BM (2018). Learning by neural reassociation. Nat Neurosci 21, 607–616. [PubMed: 29531364]

Heilbronner SR, and Hayden BY (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. Annual review of neuroscience 39, 149–170.

Herculano-Houzel S (2016). The Human Advantage: A New Understanding of How Our Brain Became Remarkable

Herry C, and Johansen JP (2014). Encoding of fear learning and memory in distributed neuronal circuits. Nat Neurosci 17, 1644–1654. [PubMed: 25413091]

Kaas JH, and Herculano-Houzel S (2017). What Makes the Human Brain Special: Key Features of Brain and Neocortex In The Physics of the Mind and Brain Disorders: Integrated Neural Circuits Supporting the Emergence of Mind, Opris I, and Casanova MF, eds. (Cham: Springer International Publishing), pp. 3–22.

Kolling N, Wittmann MK, Behrens TE, Boorman ED, Mars RB, and Rushworth MF (2016). Value, search, persistence and model updating in anterior cingulate cortex. Nat Neurosci 19, 1280–1285. [PubMed: 27669988]

Likhtik E, and Paz R (2015). Amygdala-prefrontal interactions in (mal)adaptive learning. Trends Neurosci 38, 158–166. [PubMed: 25583269]

Lin J (2006). Divergence measures based on the Shannon entropy. IEEE Trans Inf Theor 37, 145–151.

Livneh U, and Paz R (2012a). Amygdala-prefrontal synchronization underlies resistance to extinction of aversive memories. Neuron 75, 133–142. [PubMed: 22794267]

Livneh U, and Paz R (2012b). Aversive-bias and stage-selectivity in neurons of the primate amygdala during acquisition, extinction, and overnight retention. J Neurosci 32, 8598–8610. [PubMed: 22723701]

Macke JH, Berens P, Ecker AS, Tolias AS, and Bethge M (2009). Generating spike trains with specified correlation coefficients. Neural Comput 21, 397–423. [PubMed: 19196233]

MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC, Aureli F, Baker JM, Bania AE, Barnard AM, et al. (2014). The evolution of self-control. Proc Natl Acad Sci U S A 111, E2140–2148. [PubMed: 24753565]

MacLeod CE, Zilles K, Schleicher A, Rilling JK, and Gibson KR (2003). Expansion of the neocerebellum in Hominoidea. J Hum Evol 44, 401–429. [PubMed: 12727461]

Maimon G, and Assad JA (2009). Beyond Poisson: Increased Spike-Time Regularity Across Primate Parietal Cortex. Neuron 62, 426–440. [PubMed: 19447097]

Marr D (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information (Henry Holt and Co., Inc.).

Milad MR, and Quirk GJ (2012). Fear extinction as a model for translational neuroscience: ten years of progress. Annu Rev Psychol 63, 129–151. [PubMed: 22129456]

Mitchell JF, Sundberg KA, and Reynolds JH (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron 55, 131–141. [PubMed: 17610822]

Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, and Pouget A (2014). Information-limiting correlations. Nat Neurosci 17, 1410–1417. [PubMed: 25195105]

Murray EA (2007). The amygdala, reward and emotion. Trends Cogn Sci 11, 489–497. [PubMed: 17988930]

Niven JE, and Laughlin SB (2008). Energy limitation as a selective pressure on the evolution of sensory systems. The Journal of experimental biology 211, 1792–1804. [PubMed: 18490395]

Ohiorhenuan IE, Mechler F, Purpura KP, Schmid AM, Hu Q, and Victor JD (2010). Sparse coding and high-order correlations in fine-scale cortical networks. Nature 466, 617–621. [PubMed: 20601940]

Okun M, Steinmetz N, Cossell L, Iacaruso MF, Ko H, Bartho P, Moore T, Hofer SB, Mrsic-Flogel TD, Carandini M, et al. (2015). Diverse coupling of neurons to populations in sensory cortex. Nature 521, 511–515. [PubMed: 25849776]

Pare D, and Smith Y (1993). Distribution of GABA immunoreactivity in the amygdaloid complex of the cat. Neuroscience 57, 1061–1076. [PubMed: 8309543]

Pare D, Smith Y, and Pare JF (1995). Intra-amygdaloid projections of the basolateral and basomedial nuclei in the cat: Phaseolus vulgaris-leucoagglutinin anterograde tracing at the light and electron microscopic level. Neuroscience 69, 567–583. [PubMed: 8552250]

Paz R, Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, and Fried I (2010). A neural substrate in the human hippocampus for linking successive events. Proc Natl Acad Sci U S A 107, 6046–6051. [PubMed: 20231430]

Phelps EA, and LeDoux JE (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. Neuron 48, 175–187. [PubMed: 16242399]

Quirk GJ, and Beer JS (2006). Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. Curr Opin Neurobiol 16, 723–727. [PubMed: 17084617]

Quiroga RQ, Nadasdy Z, and Ben-Shaul Y (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput 16, 1661–1687. [PubMed: 15228749]

Quiroga RQ, Reddy L, Kreiman G, Koch C, and Fried I (2005). Invariant visual representation by single neurons in the human brain. Nature 435, 1102–1107. [PubMed: 15973409]

Resnik J, and Paz R (2015). Fear generalization in the primate amygdala. Nat Neurosci 18, 188–190. [PubMed: 25531573]

Rieke F, Warland D, Steveninck R.d.R.v., and Bialek W (1999). Spikes: exploring the neural code (MIT Press).

Runyan CA, Piasini E, Panzeri S, and Harvey CD (2017). Distinct timescales of population coding across cortex. Nature 548, 92–96. [PubMed: 28723889]

Sadtler PT, Quick KM, Golub MD, Chase SM, Ryu SI, Tyler-Kabara EC, Yu BM, and Batista AP (2014). Neural constraints on learning. Nature 512, 423–426. [PubMed: 25164754]

Salzman CD, and Fusi S (2010). Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. Annu Rev Neurosci 33, 173–202. [PubMed: 20331363]

Schneidman E, Berry MJ 2nd, Segev R, and Bialek W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440, 1007–1012. [PubMed: 16625187]

Shannon CE (1997). The mathematical theory of communication. 1963. MD computing : computers in medical practice 14, 306–317. [PubMed: 9230594]

Shenhav A, Cohen JD, and Botvinick MM (2016). Dorsal anterior cingulate cortex and the value of control. Nat Neurosci 19, 1286–1291. [PubMed: 27669989]

Softky W, and Koch C (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. The Journal of Neuroscience 13, 334–350. [PubMed: 8423479]

Strong SP (1998). Entropy and information in neural spike trains. Phys Rev Lett 80, 197–200.

Swanson LW, and Petrovich GD (1998). What is the amygdala? Trends Neurosci 21, 323–331. [PubMed: 9720596]

Taub AH, Perets R, Kahana E, and Paz R (2018a). Oscillations Synchronize Amygdala-to-Prefrontal Primate Circuits during Aversive Learning. Neuron 97, 291–298 e293. [PubMed: 29290553]

Taub AH, Shohat Y, and Paz R (2018b). Long time-scales in primate amygdala neurons support aversive learning. Nat Commun 9, 4460. [PubMed: 30367056]

Treves A, and Panzeri S (1995). The upward bias in measures of information derived from limited data samples. Neural Comput 7, 399–407.

Treves A, Panzeri S, Rolls ET, Booth M, and Wakeman EA (1999). Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. Neural Comput 11, 601–632. [PubMed: 10085423]

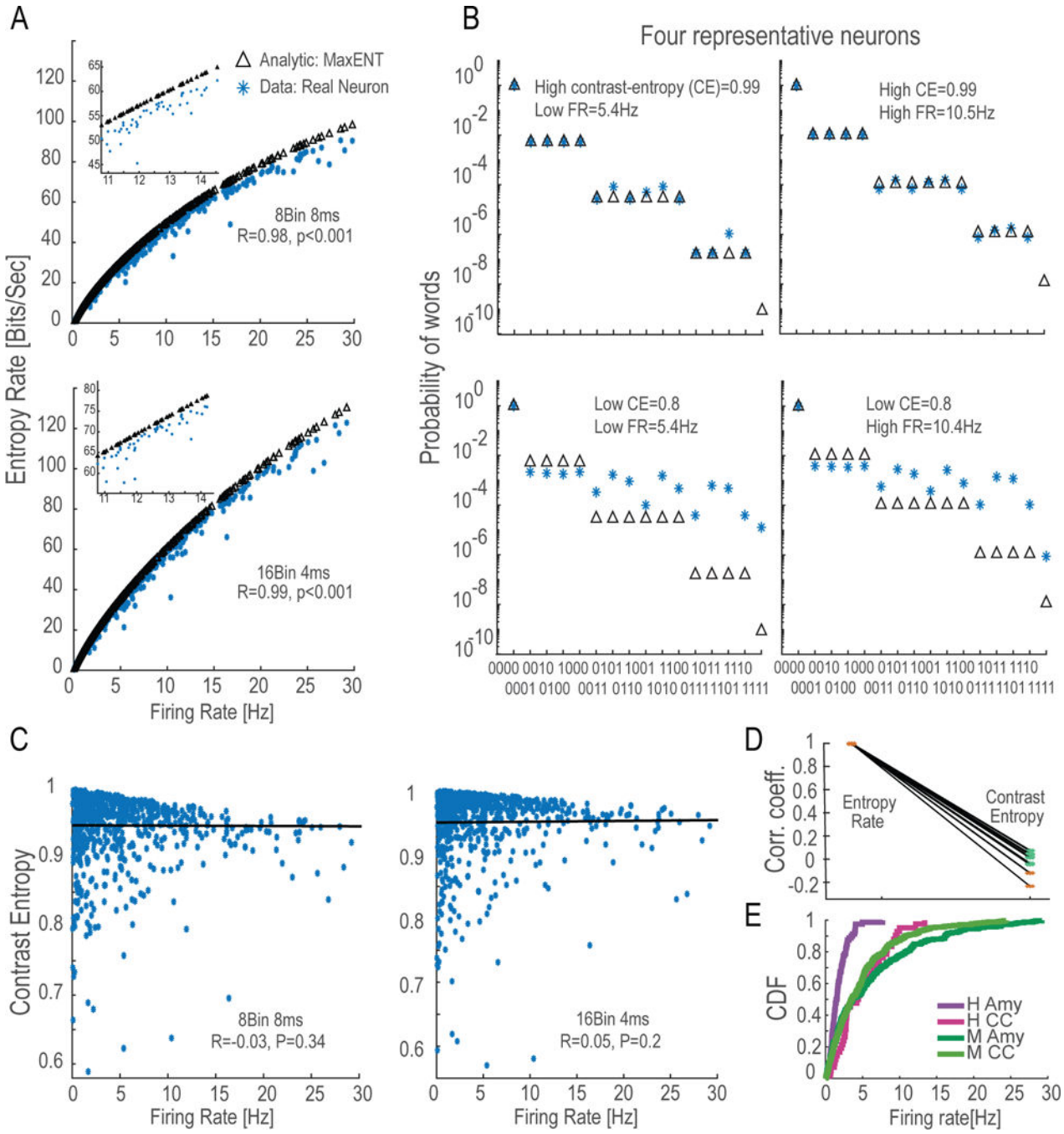

Figure 1. Contrast-entropy: efficient use of information-capacity in spike trains

A. Actual recorded neurons (blue asterisks) compared to the maximal entropy-rate for the same firing-rate (i.e. of an analytic neuron with the same FR, black triangles). Although the entropy-rate increases with firing rate (FR), the FR does not fully account for the entropy (see insets). The proportion of the entropy-rate from the analytical maximum (eq' 2) is defined as the *contrast-entropy* (eq' 3). Shown are two options of *letter* and *word* combinations (upper panel words of eight 8ms letters; lower panel words of 16 4ms letters).

B. Four examples of the difference between entropy-rate of neurons and the optimal word distribution. Shown is the probability of each word (a word of four 1ms letters in these examples), in a recorded neuron (blue asterisks) and for the analytic-optimal neuron (black triangles). Two examples of low contrast-entropy are shown (bottom row) and two of high contrast-entropy (top row), for low (left column) and high (right column) firing rates.
C. Contrast-entropy (eq' 3) measures how much neurons exploit their potential for entropy-rate given their overall FR, shown for words of eight 8ms letters (left) and words of 16 4ms letters (right).
D. Correlation between FR and contrast-entropy is significantly lower than the correlation between FR and entropy rate in all 15 letter-word combinations ($p<0.001$ for all, Fisher z-test), and all correlations between FR and contrast-entropy are close to zero (Supp.Table.1).
E. Cumulative density function (CDF) of the firing rates for both regions and species.

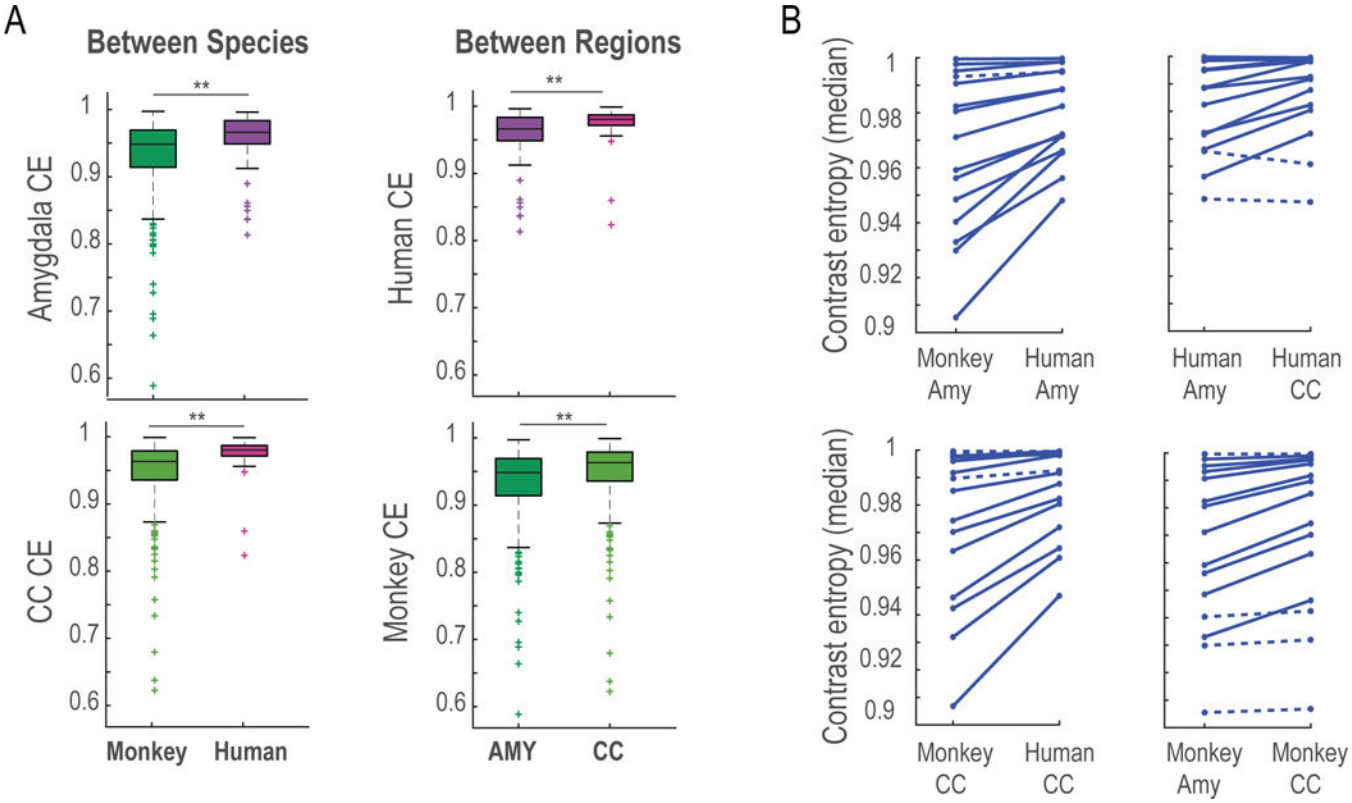

Figure 2. Differential Contrast-entropy across regions and species

A. Higher contrast-entropy is measured in humans compared to monkeys (left column), for both amygdala and cingulate-cortex (CC); and higher contrast entropy is measured in the CC compared to the amygdala (right), in both species. Shown for words of eight 8ms letters ($p<0.01$ for all, Wilcoxon signed-rank test, corrected for multiple comparisons. * $p<0.05$; ** $p<0.01$; *** p<e-3; and so forth).

B. Consistency of finding across almost all 15 word-letter combinations. Lower contrast-entropy in monkey amygdala compared to human amygdala (upper-left); Lower contrast-entropy in human amygdala compared to human CC (upper-right); Lower contrast entropy in monkey CC compared to human CC (lower-left); Lower contrast entropy in monkey amygdala compared to monkey CC (lower-right). Each panel shows all 15 letter-word combinations (significant difference in solid lines, $p<0.05$ Wilcoxon signed-rank test).

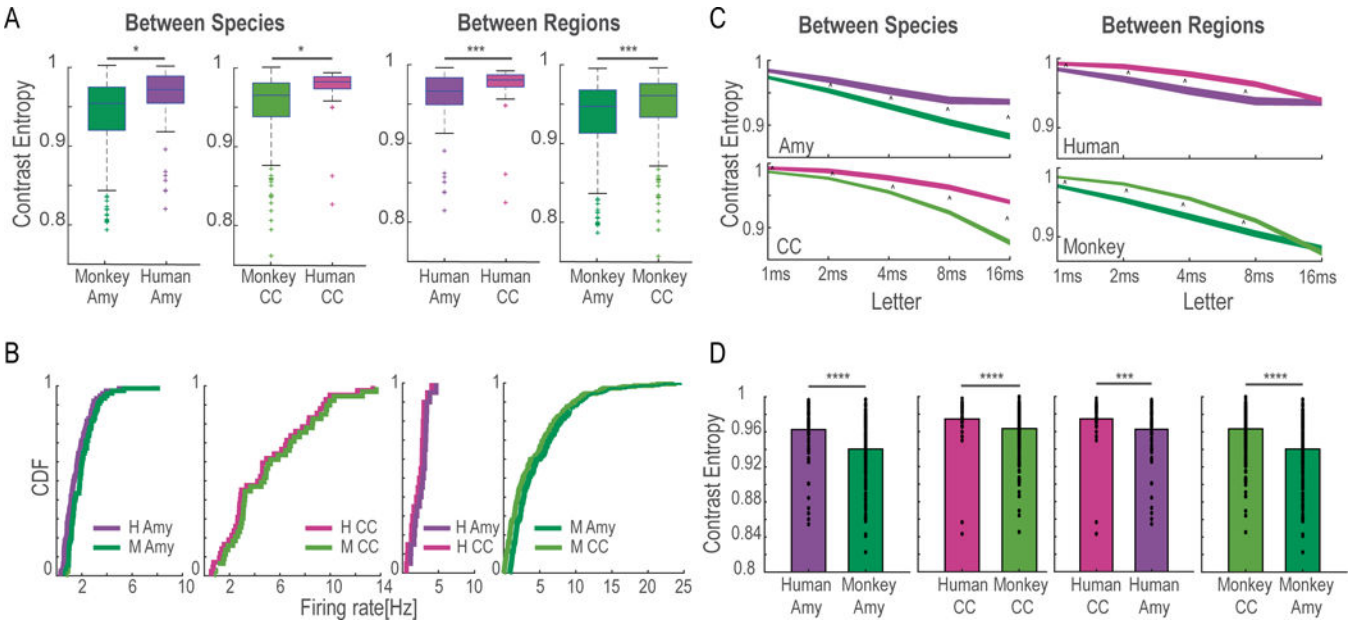

Figure 3. Differential Contrast-entropy (efficiency) is independent of firing rates

A. Matching firing-rates (see B) validates results: Lower contrast-entropy in monkey amygdala compared to human amygdala (left-most); Lower contrast-entropy in monkey cingulate-cortex (CC) compared to human CC (middle-left); Lower contrast entropy in human amygdala compared to human CC (middle-right); Lower contrast entropy in monkey amygdala compared to monkey CC (right-most). * $p<0.05$ and *** $p<0.001$ Wilcoxon signed-rank test.

B. Matching neurons with the same FR in monkey amygdala and human amygdala (left-most), and similarly for monkey and human CC (middle-left); and matching neurons with the same FR in human amygdala and human CC (middle-right), and similarly for human amygdala and CC (right-most). See methods for selection process. Shown are the cumulative distributions of FR (lines are deliberately slightly shifted for presentation only). We re-performed the analyses after equating FRs for further validation of the results from Fig.1.

C. Estimated contrast-entropy per neuron at infinite-word-length (methods). Shown is the mean +/− S.E.M (shaded) for all letter options. Upper-left: monkey amygdala is lower than human amygdala; Lower-left: monkey CC is lower than human CC; Upper-right: human amygdala is lower than human CC; Lower-right: monkey amygdala is lower than monkey CC (^ $p<0.05$, Wilcoxon signed-rank test).

D. Contrast-entropy for resampling by randomly choosing segments of 20 consecutive minutes. Shown is the mean overlaid with the individual neurons (Supp.Table.2). The order of contrast-entropy across the four paired comparisons (species and regions) is maintained.

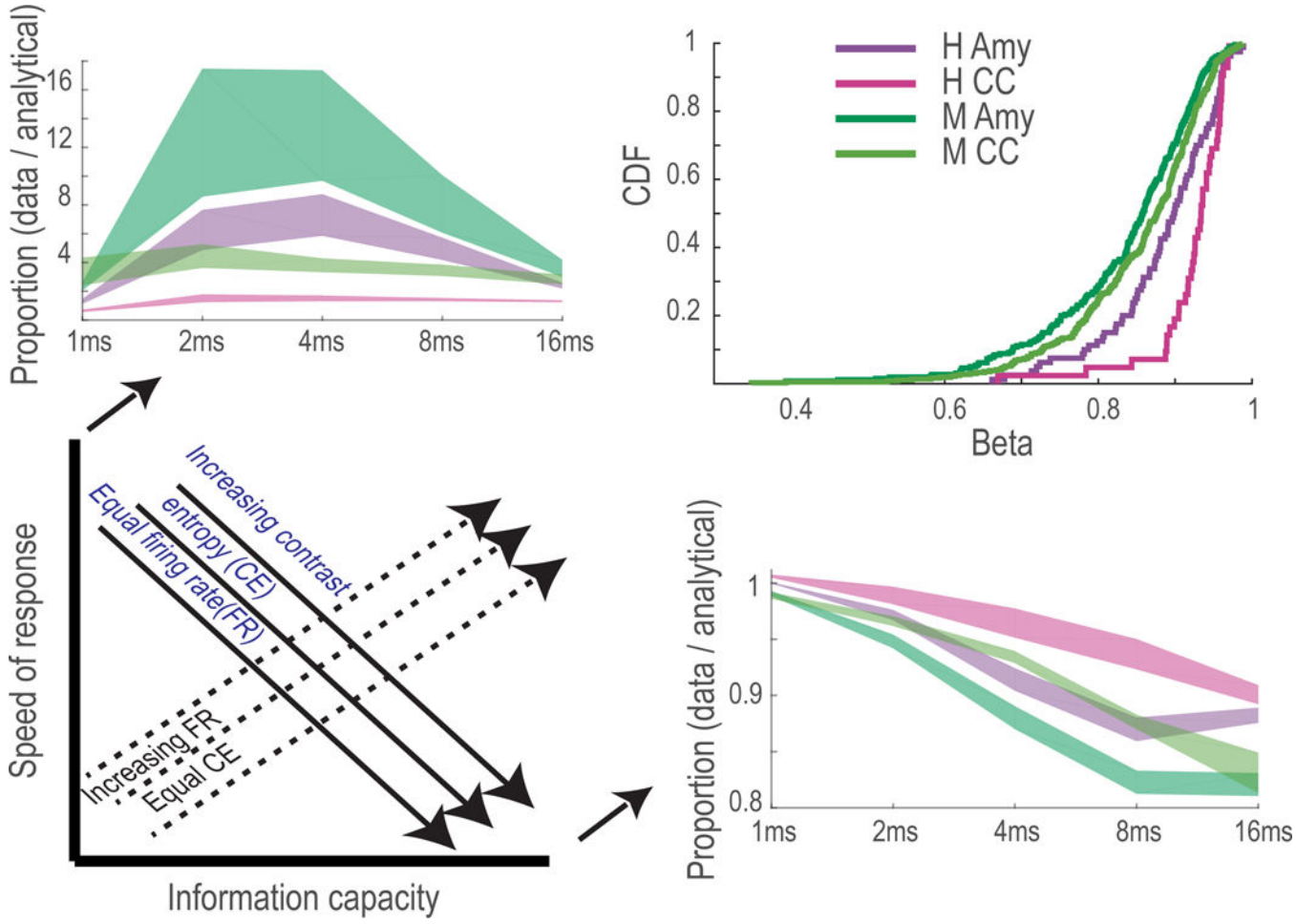

Figure 4. A tradeoff between efficiency and speed-of-response (robustness) in single-neurons.

Modelling real neurons by a two-state Markov process (methods) allows control of spike distribution for a given FR, and hence their efficiency (entropy-exploitation, β ranges from 0 to 1, minimal to maximal entropy-rate). The model was fitted for all neurons and the cumulative-density-function (CDF, upper-right) is presented as a function of β per region and species. This revealed again the order from monkey amygdala to human CC ($P<0.001$ for all, Kolmogorov-Smirnov test).

Importantly, the model unveils the tradeoff between efficiency and speed/vigor of response (robustness), where speed/vigor of response is defined as higher spike density (i.e. higher probability for words with more spikes). For a given FR, a neuron can exploit the distribution of spikes to achieve more or less entropy but at the cost of reducing robustness (lower-left, solid lines). Such tradeoff is orthogonal to increase in FR (dashed lines), which is physiologically limited.

This tradeoff is validated by the data across regions and species: the probability of words with one spike only (e.g. ‘1000’, ‘0100’) reveals the order across regions and species as seen previously for the general case when measuring contrast-entropy (lower-right, shown for all words with 4 letters). In contrast, the probability for words with more than one spike (e.g. ‘1010’, ‘1001’), unveils a reverse order (upper-left): there are more such words in the amygdala and in monkeys, indicating higher robustness. See text and methods for full

details. Shown is the mean +/− S.E.M of the proportion between the number of data neurons and analytical neurons with the same word distribution.

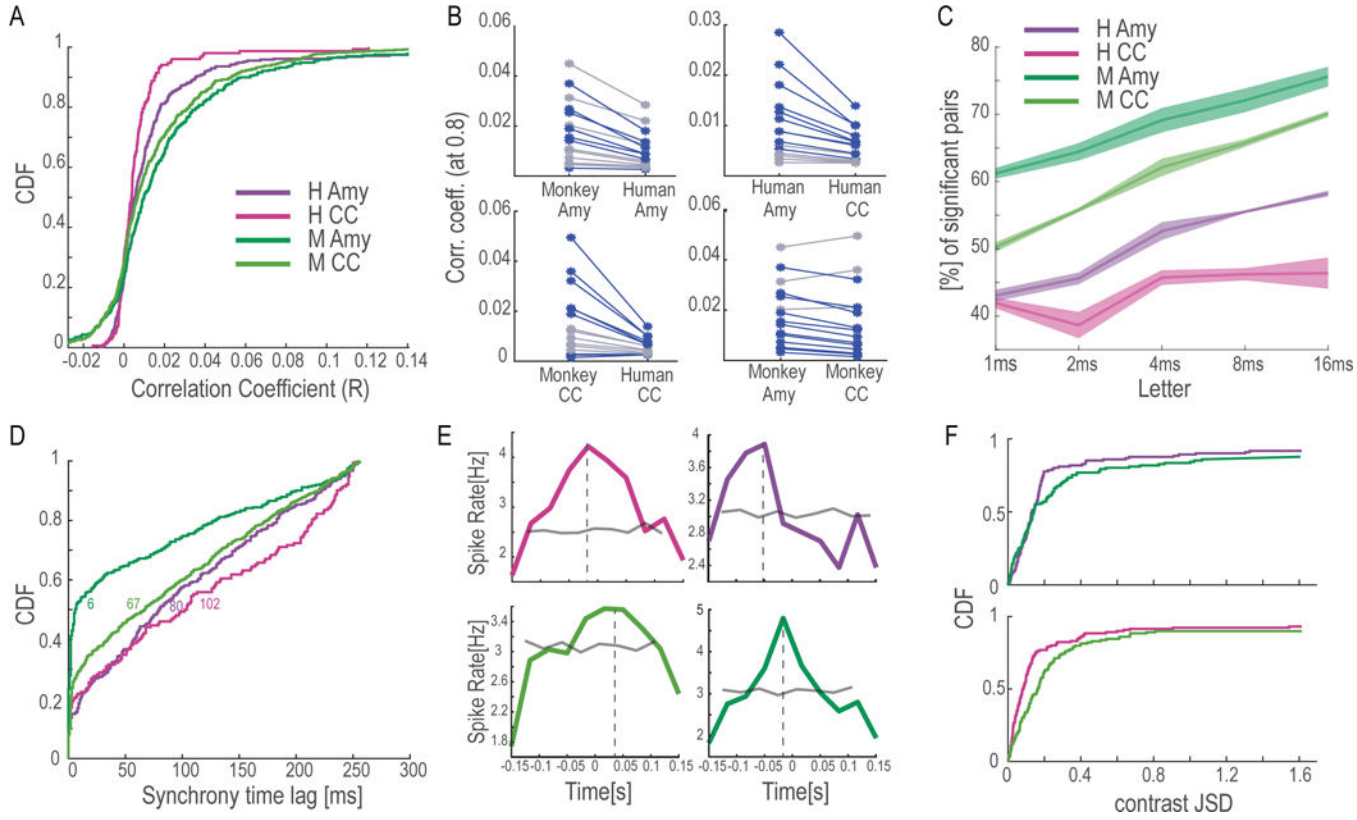

Figure 5. Robustness in pairs of neurons is higher in the amygdala and monkeys.

A. Cumulative density function of correlation coefficients in pairs of simultaneously recorded neurons. Higher values are obtained in the amygdala of both species and in monkeys ($p<0.05$ for all, Kolmogorov-Smirnov test)

B. The correlation coefficient at 0.8 of the CDF (as in A) for all 15 letter-word combinations, compared for all pairs of regions and species (Blue indicates $p<0.05$, Kolmogorov-Smirnov test). For most cases, amygdala is higher than CC in both species, and monkey is higher than human in both regions.

C. More significant pairwise correlations in the amygdala and in monkeys, for the 5 letter lengths (4 words). Shown is the percentage of significant correlations, mean+/−SEM (based on random segments of 20 min)

D. CDF of the time lag at the maximum correlation, obtained from standard pairwise cross-correlations (see E). Numbers indicate the lag at CDF=0.5. Smaller lags indicate more synchrony, leading to better downstream summation, and hence suggest robustness. There is more synchrony in monkeys and in the amygdala.

E. Four example of cross-correlations, one from each region and species. The dashed line marks the time of the maximum correlation (for each species shown is one example with a lag near zero and one further away). Gray lines represent baseline obtained from 50 circular shuffles.

F. To compare overlap in words (shared vocabulary) as another measure of robustness, we calculated the contrast-JSD (Jensen-Shannon-divergence, methods). Pairs of neurons in humans and in the CC have a JSD that is closer to the analytical JSD (for a pair with similar

firing-rates), and hence there is more overlap in the words used by pairs of neurons in monkeys and in the amygdala.

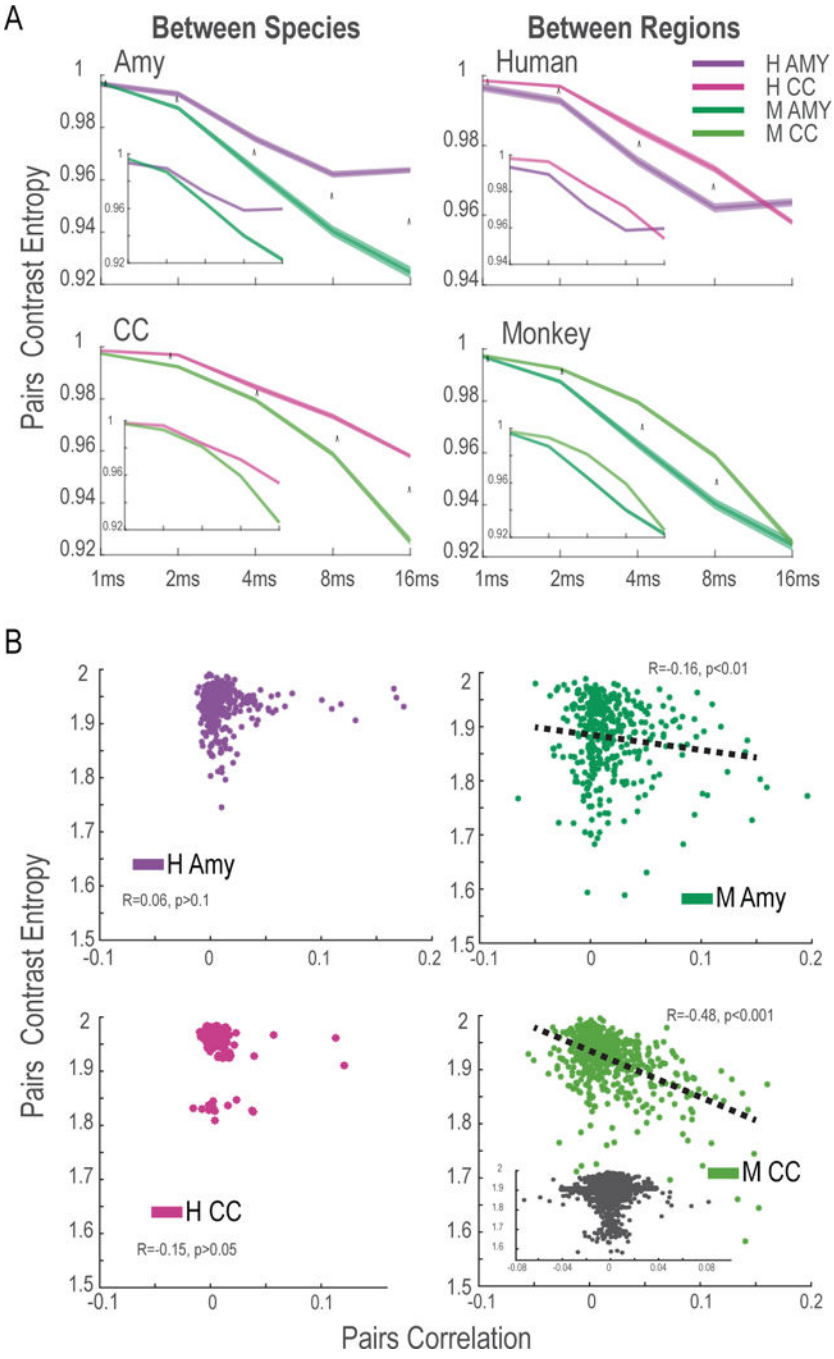

Figure 6. Efficiency vs. robustness in pairs of neurons within regions

A. Combined Contrast-entropy in pairs of simultaneously recorded neurons show, from top to bottom and right to left: higher efficiency in the human amygdala compared to monkey amygdala; higher efficiency in human CC compared to monkey CC; higher efficiency in human CC compared to human amygdala; and higher efficiency in monkey CC compared to monkey amygdala. Shown for all letter options. Inset shows the same finding for triplets of neurons that were recorded simultaneously.

B. The pairs contrast-entropy plotted against the correlation-coefficient between the same two neurons, for all four regions. No relationship was found in humans, and in contrast, an inverse significant relationship (tradeoff) was found only in monkeys ($p<0.01$, Pearson correlation). The bottom-right inset shows lack of relationship after taking pairs that were recorded in different days ($p>0.1$), demonstrating that the tradeoff is not necessary and an empirical result in monkeys.

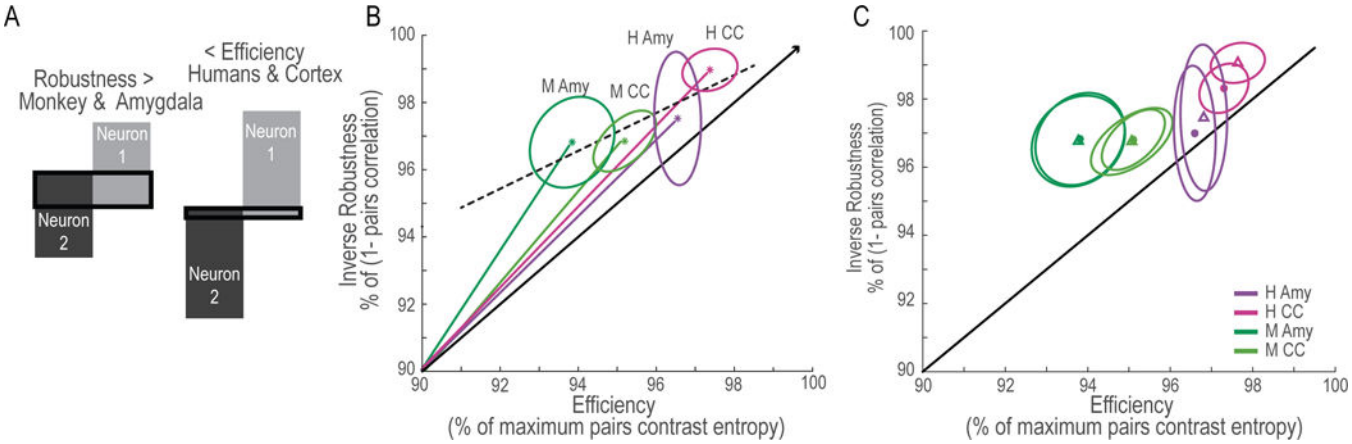

Figure 7. An efficiency-robustness tradeoff across regions and species.

A. A scheme of the differences we find between efficiency and robustness. Single units and pairs in humans and the cortex have higher entropy and more rich vocabulary (right). In contrast, neurons in monkeys and in the amygdala of both species exhibit less efficiency (left), yet exhibit higher correlations, synchrony, and overlap in words (marked in black).

B. For each region, shown is the (1-robustness) plotted against the efficiency. Both are presented as percentage of the maximum (pairs-contrast-entropy for the x-axis, and 1-correlation-coefficient for the y-axis). Plotted is the mean for each region and the error-ellipse over all neurons from this region. There is a linear relationship from monkey amygdala to human CC (dashed line, $p<0.01$). Black arrow indicates when both efficiency and inverse robustness increase at a similar rate.

C. Different behavioral paradigms produce similar tradeoff. Similar presentation as in B, but when separated into two states: neural activity taken from the period surrounding presentation of external stimuli and from the period without an external stimulus.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Propionic acid	Sigma-Aldrich, Israel	Cat# P1386; CAS ID: 79–09-4
Mineral oil	Sigma-Aldrich, Israel	Cat# M5904–500ML; CAS ID: 8042–47-5
Experimental Models: Organisms/Strains		
Adult male Macaca fascicularis	B.F.C. Monkey Breeding Farm, Israel	N/A
Humans	Patients with pharmacologically intractable Epilepsy, UCLA hospital	N/A
Software and Algorithms		
MATLAB (v.R2016b)	MathWorks	https://ch.mathworks.com/products/new_products/release2016b.html
Offline Sorter (v.3)	Plexon	https://plexon.com/products/offline-sorter/
Other		
Glass/Narylene-coated tungsten microelectrodes	Alpha Omega; We-Sense	N/A
Polyurethane probe containing platinum–iridium microwires	Costum-made, UCLA	N/A
AlphaLab SnR, Neural recording system	Alpha Omega, Israel	N/A
64-channel Neuralynx™ system	Neuralynx™	N/A