nature COMMUNICATIONS

ARTICLE

Received 9 May 2016 | Accepted 12 Oct 2016 | Published 24 Nov 2016

DOI: 10.1038/ncomms13526 OPEN

Global gain modulation generates time-dependent urgency during perceptual choice in humans

Peter R. Murphy[1,2], Evert Boonstra[1] & Sander Nieuwenhuis[1]

Decision-makers must often balance the desire to accumulate information with the costs of protracted deliberation. Optimal, reward-maximizing decision-making can require dynamic adjustment of this speed/accuracy trade-off over the course of a single decision. However, it is unclear whether humans are capable of such time-dependent adjustments. Here, we identify several signatures of time-dependency in human perceptual decision-making and highlight their possible neural source. Behavioural and model-based analyses reveal that subjects respond to deadline-induced speed pressure by lowering their criterion on accumulated perceptual evidence as the deadline approaches. In the brain, this effect is reflected in evidence-independent urgency that pushes decision-related motor preparation signals closer to a fixed threshold. Moreover, we show that global modulation of neural gain, as indexed by task-related fluctuations in pupil diameter, is a plausible biophysical mechanism for the generation of this urgency. These findings establish context-sensitive time-dependency as a critical feature of human decision-making.

[1] Institute of Psychology and Leiden Institute for Brain and Cognition, Leiden University, 2333 AK Leiden, The Netherlands. [2] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. Correspondence and requests for materials should be addressed to P.R.M. (email: murphyp7@tcd.ie).

Decision-makers are adept at trading speed for accuracy to meet contextual demands[1–3]. If time is at a premium, decisions can be made quickly at the potential expense of accuracy. Conversely, the decision-making process can also be prolonged to facilitate additional information gathering and more accurate choices. By negotiating the speed-accuracy tradeoff (SAT) in this manner, behaving agents can maximize their rate of reward in environments with different temporal constraints[4–6].

Studies of decision-making support a broad class of models in which noisy evidence for each available choice is accumulated over time and a decision is made once the accrued evidence passes a criterial level, termed the decision bound[7–13]. Within this framework, one intuitive and parsimonious account of SAT asserts that speed emphasis is regulated by adjusting the level of the decision bound, such that less evidence is required for decision commitment in situations that demand faster decision-making. Aside from such situational or 'static' adjustments, the bound is typically assumed to be constant over the course of a single decision, thereby enforcing a fixed policy on commitment for a given decision-making context. For decades, models that invoked such a context-dependent, time-invariant bound have provided good fits to empirical SAT data (for example, refs 9,14–17).

Recently, convergent lines of research have brought the principle of time-invariance into question. Theoretical treatments have shown that a time-invariant decision policy is sub-optimal when the potential cost of continued deliberation grows over time[18,19]—as is the case, for example, when speed pressure is generated by means of a temporal deadline on choices[20]. In such settings, maximizing reward instead relies on dynamically lowering the evidence required for commitment as elapsed decision time increases. Additionally, recent primate single-unit recording studies indicate that a time-dependent, evidence-independent influence on the decision process is observable in the activity of neurons that reflect evolving decision formation[21–23], and moreover, that the strength of this time-dependency is highly sensitive to SAT manipulations. In particular, in both lateral intraparietal[22] and dorsal premotor[23] neurons, greater speed emphasis manifests in a combination of statically increased baseline firing rates (see also ref. 24) and a clear evidence-independent increase in firing rates with greater elapsed time. It has been proposed that these contextually-sensitive influences combine to form a neural urgency signal that expedites the evolving decision process by driving it closer to a fixed threshold, which translates to a dynamic criterion on evidence[19,22,23,25–28].

While these findings have illuminated the mechanistic basis of SAT regulation in non-human primates, time-invariance remains a dominant assumption in the human decision-making literature[10] and recent empirical and model comparison reports have reinforced this stance[15,29–31]. Moreover, even in non-human primates, little is known about the neurophysiological source of urgency. In the present study, we address these outstanding issues. We first present convergent behavioural, electrophysiological and model-based evidence that human subjects do invoke an urgency signal with both static and time-dependent components to adapt to deadline-induced speed pressure. Next, we show that global modulation of neural gain, as reflected in task-related fluctuations in pupil diameter, is a plausible biophysical mechanism for the generation of urgency. Lastly, we report that human behaviour bears hallmarks of time-dependency even when speed pressure is mild and not a central feature of task design.

Results

Behaviour under deadline and free response. In the first experiment that we report, twenty-one individuals made two-alternative perceptual decisions about the dominant direction of motion of a cloud of moving dots[32] (Fig. 1a). Each subject performed this task at a single level of discrimination difficulty that was tailored to their perceptual threshold, but under two levels of speed emphasis. In the 'free response' (FR) regime, subjects were under no external speed pressure, were instructed to be as accurate as possible, and were monetarily rewarded (penalized) for correct (incorrect) decisions. In the 'deadline' (DL) regime, the same task instructions and incentive scheme applied, with the addition of an especially heavy penalty—ten times that for an incorrect decision—if a decision was not made by 1.4 s after motion onset. The speed pressure imposed by this deadline led to faster median response times (RTs: DL = 0.70 ± 0.02 s; FR = 1.19 ± 0.07 s; $t_{20} = 8.1$, $P < 1 \times 10^{-6}$) and less accurate decision-making (DL = 77.8 ± 1.2%; FR = 86.8 ± 1.3%; $t_{20} = 6.6$, $P < 1 \times 10^{-5}$) relative to the FR regime (Fig. 1b).

Given the large penalty for missed deadlines, a sensible strategy in the DL regime is to always execute a response before the deadline[20]. Indeed, subjects missed the deadline on a median of only 0.14 ± 0.13% of trials, compared to 38.8 ± 4.0% of RTs

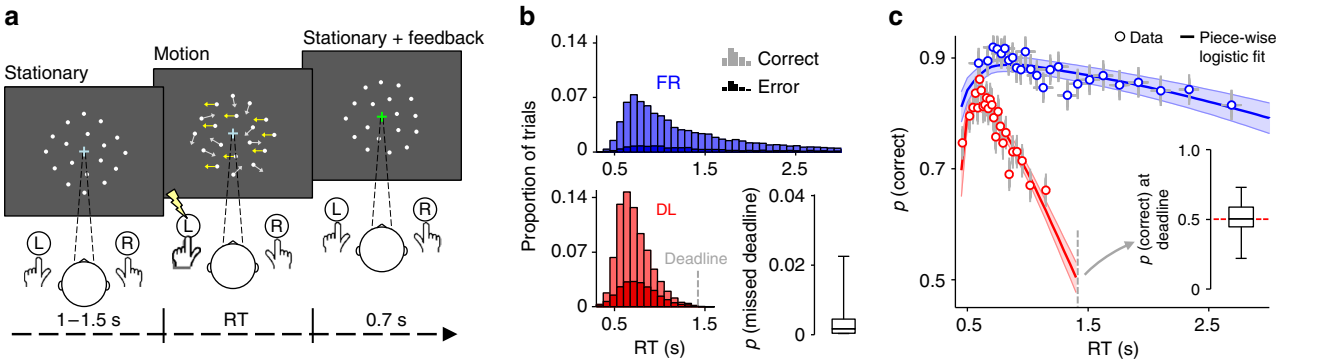

Figure 1 | Perceptual task and associated behaviour. (**a**) Schematic of a single-trial of the random dot motion task. (**b**) Subjects performed under 'free response' (FR) and 'deadline' (DL) conditions to manipulate speed pressure. Histograms depict pooled RT distributions from each condition. Box plot at lower right shows the sample median (centre line), interquartile range (box) and full range (whiskers) of proportion of missed deadlines in the DL condition. (**c**) Conditional accuracy functions. Points indicate mean accuracy of trials sorted by RT into 25 equal-sized bins and coloured lines show best fits of piece-wise logistic regressions to each subject's single-trial data. Error bars and shaded areas indicate ± s.e.m. of data points and regression lines, respectively. Box plot at inset shows the sample median (centre line), interquartile range (box) and full range (whiskers) of estimated accuracy at deadline in the DL condition.

exceeding this time in the FR regime (Fig. 1b). In accumulation-to-bound models of decision-making, this marked change in behaviour can primarily be achieved in two ways: by imposing a diminishing criterion on accumulated evidence as the deadline draws nearer, culminating in zero required evidence (and consequently, chance performance) around the time of the deadline; or, by lowering an otherwise static criterion sufficiently to ensure that effectively all decisions are made before the deadline, but are always based on the same quantity of accumulated evidence. While the latter mechanism predicts that the slope of the conditional accuracy function (CAF) relating accuracy to RT will be similar across speed emphasis regimes and that performance will generally not reach chance levels, the former predicts that the slope of the CAF will be substantially more negative in the DL regime and should arrive at approximately chance performance by the time of the deadline. Thus, empirical CAFs can, in principle, be used to arbitrate between different mechanistic accounts of SAT adjustment.

We employed single-trial logistic regression to estimate the shape of the empirical CAFs (see Methods; Fig. 1c). After accounting for a small percentage of fast inaccurate decisions, the estimated CAF slopes were negative in both the FR ($\beta = -0.35 \pm 0.07$, $t_{20} = -5.2$, $P < 1 \times 10^{-4}$) and DL ($\beta = -2.03 \pm 0.22$, $t_{20} = -9.3$, $P < 1 \times 10^{-8}$) regimes, but much more so in the latter (FR versus DL: $t_{20} = -8.3$, $P < 1 \times 10^{-7}$). Moreover, using the DL regression fits to estimate accuracy at the time of the deadline revealed that this was not different from chance across subjects ($50.4 \pm 2.7\%$; one-sample t-test with $H_0 = 50\%$: $t_{20} = 0.2$, $P = 0.9$).

A negative CAF slope by itself does not necessarily imply time-dependency in the decision process; indeed, it should be expected whenever the strength of decision evidence fluctuates across trials, because trials with weak evidence will tend to be both slower and less accurate and thereby produce an asymmetry in correct and incorrect RT distributions[33]. Such evidence fluctuations can be due to variation in objective stimulus strength, but also to endogenous variation in attention or arousal[34]. We therefore examined the possibility that the CAF difference that we observed between the DL and FR conditions was simply caused by condition-related differences in arousal state. To do so, a second cohort of subjects performed the same motion discrimination task and we compared their CAFs on subsets of DL and FR trials that were precisely matched for pre-motion pupil size, a commonly-used metric of arousal and 'brain state' (see below). Even in this case of matched pupil-linked arousal, we observed a much more negative CAF slope under deadline ($t_{22} = -7.7$, $P < 1 \times 10^{-6}$; Supplementary Fig. 1).

Combined, the above observations are consistent with the adoption of a time-dependent decision policy in the DL regime. Two additional observations illuminate the nature of this policy

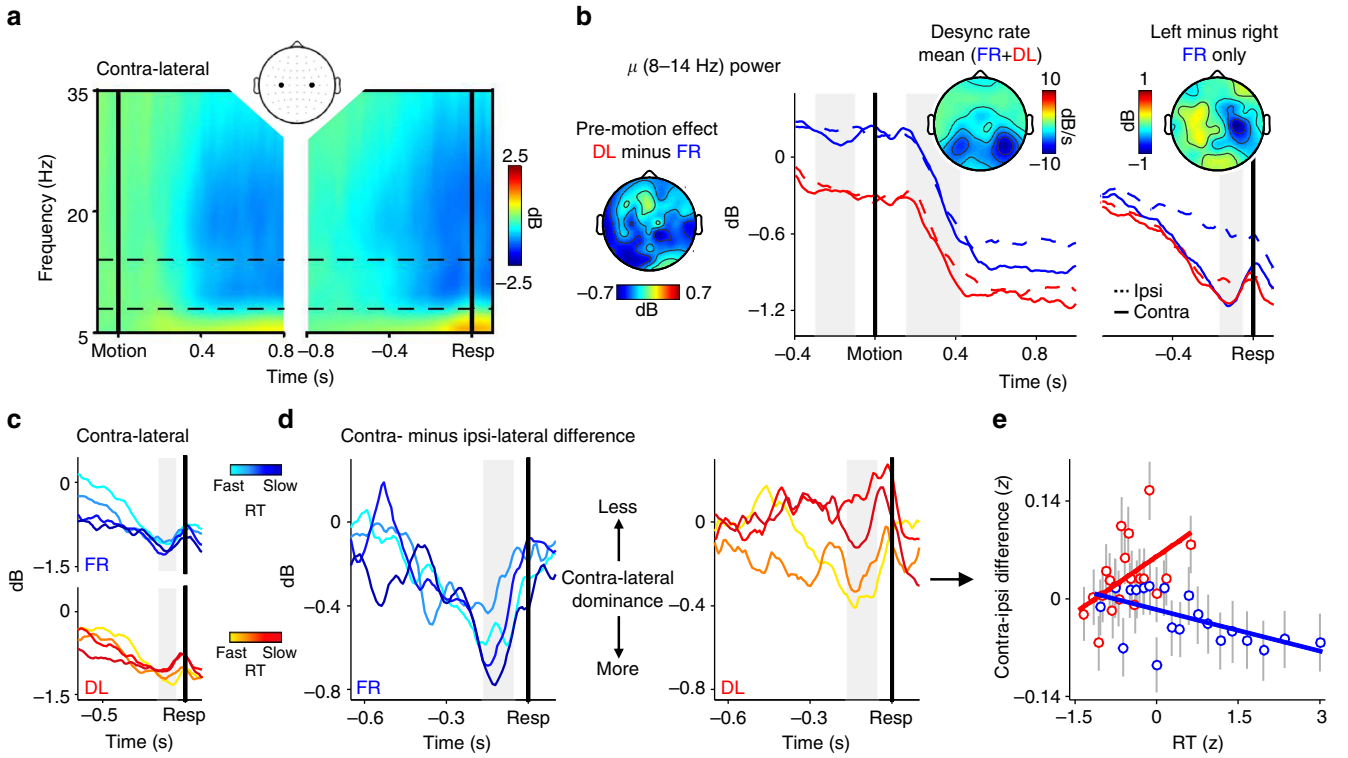

Figure 2 | μ power tracks motor preparation and reveals urgency signatures under deadline-induced speed pressure. (**a**) Time-frequency plot of oscillatory power over lateral motor channels, aligned to motion onset (left) and response (right). Plots show the trial-averaged change in power over the channel contra-lateral to the executed response on each trial, relative to a pre-motion baseline. (**b**) Onset and response-aligned μ (8–14 Hz) signals, separated by speed regime and lateralization relative to the executed response. Topographies at left, middle and right depict distribution of pre-motion effect of speed emphasis, stereotyped onset-evoked power decrease maximal over occipital scalp, and lateralization of μ power immediately prior to response execution in the FR condition, respectively. The level of contra-lateral desynchronization prior to response is highly similar across speed regimes, consistent with a common motor threshold. (**c**) Response-aligned μ signals contra-lateral to the executed response after sorting trials by RT into 4 equal-sized bins, separately for the FR and DL conditions. (**d**) Contra- minus ipsi-lateral difference waveforms, again after RT-sorting into 4 bins. Contra-lateral dominance prior to response execution decreases with slower RTs under deadline. (**e**) Scatterplot illustrating the linear relationships between RT and the contra-/ipsi-lateral μ difference for each speed regime. Points and error bars are mean ± s.e.m. of data that were z-scored within subjects, pooled across subjects and grouped into 20 bins; z-scoring was carried out across speed regimes to preserve main effects of speed emphasis. In **b**–**d**, shaded grey regions show measurement windows for scalp topographies and associated effects reported in text.

adjustment further. First, the smooth gradient of the right tail of the DL RT distributions and their associated CAF indicate that the time-dependent change in required evidence was gradual, not abrupt. Second, despite task difficulty being fixed across both speed emphasis regimes, peak decision accuracy across all RTs (as measured at the inflection point of the estimated CAFs) was reliably lower under deadline (84.3 ± 1.4 versus $89.8 \pm 1.2\%$; $t_{20} = -4.0$, $P = 0.0008$). This suggests that, further to the time-dependent effect, additional speed emphasis was generated by a static, time-invariant lowering of the criterion on accumulated evidence which ensured that even fast decisions were less accurate in the DL regime.

EEG motor preparation signatures of urgency. Although these behavioural findings suggest that greater speed emphasis under deadline was achieved by a combination of static and time-dependent adjustments to the decision policy, they are not decisive about the mechanistic basis of these adjustments. One possibility is that the decision bound itself varies with speed emphasis and progressively collapses as the deadline approaches (for example ref. 35). Alternatively, the bound might remain fixed and an urgency signal could generate speed emphasis by providing additional input to each evidence accumulator. In an attempt to adjudicate between these competing mechanistic accounts of the static and time-variant influences on decision-making behaviour, we measured scalp EEG and examined motor preparatory activity via oscillatory power in the μ (8–14 Hz) frequency range. Previous studies have shown that the commonly observed decrease in μ power during decision formation reflects dynamic motor preparation that appears to be driven by the evidence accumulation process[36–38]. Here, we used effector-selective μ signals that were contra- and ipsi-lateral to the executed response as proxies for trial-by-trial preparatory activity in favour of the chosen and unchosen motion directions, respectively.

We observed that bi-lateral motor μ signals (Fig. 2a) were sensitive to speed emphasis in several distinct ways. First, there was a reliable effect of speed emphasis on μ power prior to motion onset such that pre-motion power was lower in the DL regime compared with the FR regime (mean $\beta = -0.102 \pm 0.044$, $t_{20} = -2.3$, $P = 0.03$; Fig. 2b, left). There was no main effect of lateralization (contra- versus ipsi-) during this period ($\beta = -0.019 \pm 0.015$, $t_{20} = -1.2$, $P = 0.2$), and no speed emphasis by lateralization interaction ($\beta = 0.016 \pm 0.021$, $t_{20} = 0.8$, $P = 0.5$). Thus, potentially indicative of a static urgency effect, greater speed pressure was accompanied by increased baseline motor preparation in both effectors. Topographic visualization of this pre-motion effect revealed that although foci of decreased power in the DL regime were apparent over bi-lateral motor channels, the effect also extended over posterior scalp. However, the effect over lateral motor areas remained marginally significant even when posterior 8–14 Hz activity was included as a co-variate ($\beta = -0.068 \pm 0.038$, $t_{20} = -1.8$, $P = 0.09$), suggesting that this motor effect was at least partially distinct from the more posterior effect.

Next, we turned to μ power prior to response execution in order to examine effector-specific motor preparation at decision commitment (see Methods for rationale behind selectively focusing on this measurement period). Consistent with previous findings[36–38], there was a lateralization in pre-response μ power: a greater decrease was present in contra-lateral rather than ipsi-lateral channels, reflecting greater motor build-up in favour of the ultimately executed response (Fig. 2b, right). Accordingly, a main effect of lateralization was observed in a statistical model with lateralization and speed emphasis regime as factors ($\beta = -0.120 \pm 0.044$, $t_{20} = -2.7$, $P = 0.013$). However, this effect was also accompanied by a main effect of speed emphasis ($\beta = -0.089 \pm 0.034$, $t_{20} = -2.6$, $P = 0.016$) and a lateralization by speed emphasis interaction ($\beta = 0.095 \pm 0.025$, $t_{20} = 3.8$, $P = 0.001$). *Post-hoc* models revealed that while the expected contra/ipsi lateralization was clearly apparent in the FR regime ($\beta = -0.129 \pm 0.043$, $t_{20} = -3.0$, $P = 0.007$), it was not reliable under deadline ($\beta = -0.040 \pm 0.034$, $t_{20} = -1.2$, $P = 0.3$). Moreover, the pre-response ipsi-lateral signals representing motor preparation for the unchosen alternative were of significantly lower power in the DL relative to the FR regime ($\beta = -0.093 \pm 0.031$, $t_{20} = -3.0$, $P = 0.008$), whereas the contra-lateral signals reached a highly similar level ($P = 0.9$). All of these effects were also present when only subsets of RT-matched trials from each speed emphasis condition were analysed (Supplementary Fig. 2).

The stereotyped level of pre-response contra-lateral μ power suggests that the level of motor preparation required to execute a response was the same across both speed emphasis regimes. We also observed that this metric was invariant to RT within each regime (FR: $\beta = -0.012 \pm 0.014$, $t_{20} = -0.8$, $P = 0.4$; DL: $\beta = 0.025 \pm 0.014$, $t_{20} = 1.8$, $P = 0.09$; Fig. 2c). To the extent that pre-response μ may provide a proxy for the level of the decision bound, this pattern of findings is consistent with a fixed bound across speed emphasis regimes and decision times and therefore argues against the notion that speed emphasis is generated by bound adjustment. Instead, the lower μ power that was evident in the DL regime during the pre-motion period, and in the ipsi-lateral signal at the time of commitment, might plausibly reflect an urgency signal that provides an additional source of input to the evidence accumulation process.

In the above respects, our findings are consistent with recent studies of the neural basis of SAT regulation in non-human primates[22,23]. In a further analysis, we investigated whether, as in these studies, there was a time-dependent component to the urgency signal. In non-human primates, time-dependency in the neural urgency signal manifests as a building, common increase in the firing rates of neurons reflecting evidence accumulation for both the chosen and unchosen task alternatives[21–23]. In the case of our pre-response motor preparation signals, this time-dependent increase in common activation (or put differently, the time-dependent decrease in the difference between accumulators) should translate into a diminishing contra/ipsi lateralization with increasing RT (see Methods). Accordingly, we observed a speed regime by RT interaction ($\beta = 0.055 \pm 0.023$, $t_{20} = 2.4$, $P = 0.026$; Fig. 2d,e) in a model that examined the effect of these factors on pre-response μ lateralization. *Post-hoc* tests indicated that although there was no reliable relationship between μ lateralization and RT in the FR regime ($\beta = -0.020 \pm 0.013$, $t_{20} = -1.6$, $P = 0.1$), the strength of lateralization decreased as predicted for slower RTs in the DL regime ($\beta = 0.020 \pm 0.009$, $t_{20} = 2.2$, $P = 0.038$). This finding supports the hypothesis that, in addition to the static pre-motion effect described earlier (Fig. 2b, left), greater speed emphasis under deadline was generated by a time-dependent urgency signal that increased in magnitude as the deadline drew nearer.

Drift diffusion modelling corroborates urgency account. In light of this combined behavioural and electrophysiological support for static and time-varying urgency as mechanisms for generating greater speed emphasis, we proceeded to verify that a computational model that incorporates these features can account for the observed behavioural data. In our model, a decision is made when one of two anti-correlated evidence accumulators reaches a fixed decision bound, and the accumulators are subject to the same additive, time-varying urgency signal that is free to

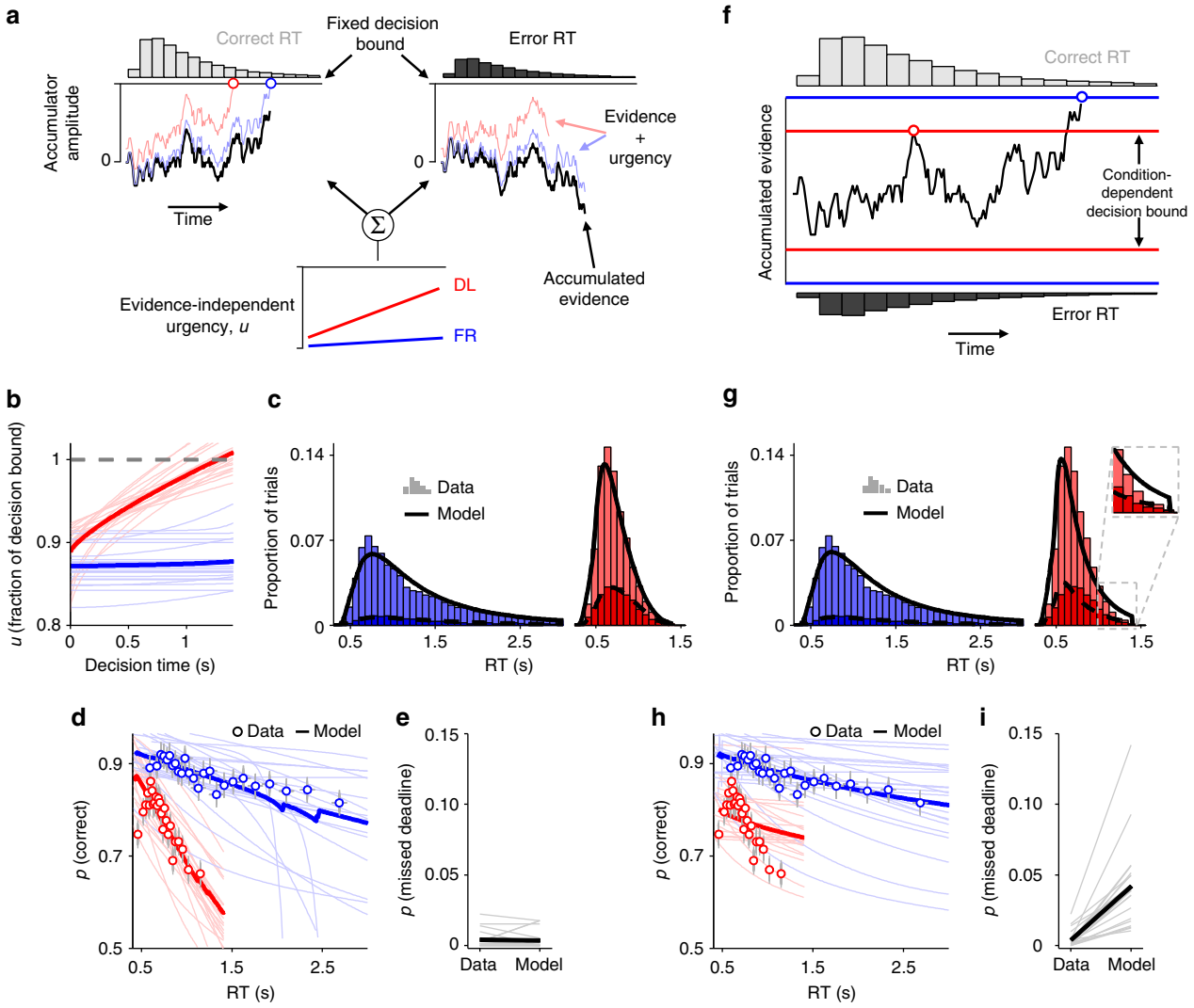

Figure 3 | Comparison of diffusion model fits with and without urgency. (**a**) Schematic representation of the drift diffusion model (DDM) with urgency and fixed decision bounds. (**b**) Urgency signals, modelled as a logistic function, derived from fits of the urgency DDM to behaviour. Lighter lines indicate fits for individual subjects; darker lines indicate group-averages. (**c**) Observed and fitted RT distributions (histograms and lines, respectively), pooled across subjects. (**d**) Conditional accuracy functions from urgency DDM fits. Points indicate mean accuracy of trials sorted by RT into 25 equal-sized bins (error bars indicate ± s.e.m.); dark/light lines depict group-average/single-subject model fits, respectively. (**e**) Observed and fitted proportion of missed deadlines. (**f**) Schematic of the regular DDM without urgency. (**g**–**i**) Fitted RT distributions (**g**), conditional accuracy functions (**h**) and proportion of missed deadlines (**i**) derived from fits of the regular DDM to behaviour.

take different shapes across speed emphasis regimes (Fig. 3a; Methods). Without urgency, this model reduces to the popular drift diffusion model (DDM) in which a single accumulation process plays out between two opposing bounds[10] (Fig. 3f).

We fit a variety of models with urgency to the data and found that the best-fitting model allowed both the shape of the urgency signal and a non-decision time parameter to vary across speed emphasis regimes (Supplementary Table 1). The rate of evidence accumulation, known as the drift rate, was allowed to vary across trials in this model (Wilcoxon signed-rank tests on BIC differences: $P<0.1$ for all pair-wise comparisons of otherwise identical model variants with and without drift rate variability), but mean drift rate and the magnitude of this between-trial variability were fixed across speed regimes. This best-fitting urgency model fit the observed RT distributions well (Fig. 3c) and was able to reproduce the key qualitative behavioural effects of increased speed emphasis under deadline (Fig. 3d,e): the more negative CAF slopes; the very low proportion of missed deadlines; and, in most subjects, the tendency toward near-chance performance at the time of the deadline. Notably, there were effects of speed emphasis on both the baseline offset and the time-varying shape of the fitted urgency signals, corresponding to greater static and time-dependent urgency under deadline, respectively, and the shape of the fitted urgency signals was highly consistent across subjects (Fig. 3b). Moreover, non-decision times were found to be marginally faster in the DL regime compared with the FR regime (Supplementary Table 2).

By contrast, the standard DDM with condition-dependent but time-invariant decision bounds (Fig. 3f) provided a considerably poorer fit to the observed data (Fig. 3g–i; Supplementary Table 1; Wilcoxon signed-rank tests on BIC differences: $P<0.001$ for all pair-wise comparisons of urgency models with their standard DDM counterparts). This poor fit stems from the fact that the standard DDM is incapable of generating a more negative CAF slope, to the extent required here, without also increasing the proportion of missed deadlines. Its main mechanism for lowering the slope of the CAF is to increase the between-trial variability in drift rate[33,34]; but, this produces a relative increase in the

proportion of trials that have a near-zero drift rate, which are less likely to reach the decision bound before the deadline. As a consequence, in our standard DDM fits, leaving between-trial variability in drift rate free to vary across speed emphasis regimes did not even yield an increase in goodness-of-fit (Supplementary Table 1). Thus, quantitative model comparisons corroborated the presence of urgency with time-dependency in the decision process under deadline.

Using the closed-form function for the urgency signal in the above model fits (equation 4), we also approximated the optimal, reward-maximizing shape of time-dependent urgency on our task for a representative set of remaining model parameters (Supplementary Fig. 3). This optimal urgency signal required a fast transition from a flat early portion to a steep deflection toward the decision bound closer to the deadline (cf. refs 20,30) that is qualitatively very different from the gradual, approximately linear urgency signals that subjects in the current study appeared to implement. As such, although our subjects responded to deadline-induced speed pressure by adjusting their decision policies in a time-dependent fashion, they failed to do so optimally. Interestingly, when the urgency signal was further constrained to be strictly linear in the optimality calculations, its reward-maximizing trajectory was matched much better by the fitted signals derived from the observed data (Supplementary Fig. 3). Combined, these observations may point to limitations of the neural mechanisms responsible for urgency generation (see Discussion).

Pupillometry highlights gain modulation as source of urgency. While the above findings describe the effects of urgency on behaviour and cortical signatures of decision-related motor preparation, they do not shed light on the neural origins of the urgency signal. Theoretical accounts have identified gain modulation, which affects the responsivity of both excitatory and inhibitory neural connections, as a potential mechanism for generating urgency in the brain[19,26,27,39,40]. However, this possibility has not been tested empirically. In the second experiment, we investigated whether global, brain-wide gain modulation, as indexed by pupil diameter, may be implicated in the injection of urgency into the decision process. Under constant luminance, changes in pupil diameter have been linked to the activity of diffusely-projecting neuromodulatory systems, in particular the locus coeruleus-noradrenergic (LC–NA) system[41–43], that are thought to control global neural gain[44–47].

A second cohort of twenty-three subjects (whose CAFs are already reported in Supplementary Fig. 1) performed the motion discrimination task optimized for measurement of decision-related changes in pupil diameter. We first examined the effect of speed emphasis on unbaselined pupil diameter prior to motion onset, which has previously been used as a proxy for 'tonic' fluctuations in neural gain[45,48]. Consistent with a static increase in gain under greater speed pressure, this metric was larger in the DL regime than in the FR regime ($t_{22} = 6.9$, $P < 1 \times 10^{-6}$; Fig. 4a).

Next, we examined the effect of speed emphasis on evoked, 'phasic' pupil dilations after motion onset and whether this effect interacted with RT, as expected of an urgency signal with a strength that depends on elapsed decision time. We observed a main effect of speed regime on trial-by-trial pupil dilation magnitude, driven by larger dilations in the DL regime than the FR regime ($\beta = 0.143 \pm 0.048$, $t_{22} = 3.0$, $P = 0.007$; Fig. 4b). Moreover, there was a significant speed regime by RT interaction ($\beta = 0.189 \pm 0.040$, $t_{22} = 4.7$, $P < 1 \times 10^{-4}$; Fig. 4c,d). *Post-hoc* tests revealed that while no reliable relationship existed between dilation magnitude and RT in the FR regime ($\beta = -0.022 \pm 0.020$, $t_{22} = -1.1$, $P = 0.3$), pupil dilations were larger for slower RTs in the DL regime ($\beta = 0.085 \pm 0.017$, $t_{22} = 5.0$, $P < 1 \times 10^{-4}$). These effects were present across a broad range of both stimulus- and response-aligned measurement windows (Supplementary Fig. 4).

We next sought to identify the most likely shape of the neural input to the pupil system during decision formation by combining linear systems analysis with formal model selection. In accordance with recent reports[49,50], the trial-related input to the pupil system was modelled as a linear superposition of three temporal components: a transient at motion onset, a transient at response, and a sustained component throughout the intervening period of decision formation, each convolved with a pupil impulse response function[51]. Using this approach, we then compared the goodness-of-fit of a variety of models in which the shape of the sustained decisional component varied (Fig. 5a; Methods). The model that best fit the pupil data from the FR regime was one in which the input to the pupil system maintained a constant amplitude throughout the decisional period (a 'boxcar'), irrespective of how long the decision took to be made (Fig. 5b). In contrast, the best fit to the DL data was provided by a model in which input strength ramped up monotonically with elapsed decision time (Fig. 5c). The latter finding reflects a truly time-dependent increase in the neural input to the pupil system in the DL regime, thus supporting the hypothesis that the gain of neural processing increased with elapsed time under speed pressure. Additionally, the boxcar and linear up-ramp remained the best-fitting models of the FR and DL data, respectively, across a wide range of different parameterizations of the pupil impulse response function (Supplementary Fig. 5).

In both speed emphasis regimes, each of the three modelled temporal components contributed significantly to the measured pupil time series (Fig. 5d,e). Hence, we also tested whether the observed relationship between pupil dilation and RT in the DL regime (Fig. 4c,d) was fully captured by the RT modulation inherent in the ramping decisional component of the associated best-fitting model, or if the onset and response components also contributed to this effect. In model variants that included additional terms representing the parametric modulation of each temporal component by RT, neither modulated term for the onset or response components contributed consistently to the DL pupil time series (Effect size for parametrically modulated onset term $= -4.5 \pm 3.2$, $t_{22} = -1.4$, $P = 0.2$; Effect size for parametrically modulated response term $= 2.7 \pm 3.2$, $t_{22} = 0.8$, $P = 0.4$). This suggests that the dilation/RT relationship in the raw DL data reflects a time-dependent modulation of input to the pupil system that was specific to the period of decision formation.

Global gain modulation alone produces urgency effects. To build on these pupillometric observations, we next verified that a combination of static and time-dependent changes in global gain is capable of producing the qualitative effects of deadline-induced speed pressure on both overt behaviour and decision-related neural dynamics. We modelled global gain modulation as a change in the slope of the input-to-output transfer function of a simple neural network that incorporates basic principles of neural computation[12] (Fig. 6a,b). Informed by our pupillometric results, gain was fixed at a low level throughout a trial in the FR regime but subject to static and time-dependent increases in the DL regime (Fig. 6c). All other model parameters, aside from non-decision time, were fixed across regimes (Methods).

When fit to the pooled behaviour of the cohort of subjects from the first experiment reported above, this model successfully reproduced all of the key qualitative effects of deadline-induced speed pressure on both behaviour (Fig. 6d,e; proportion of missed

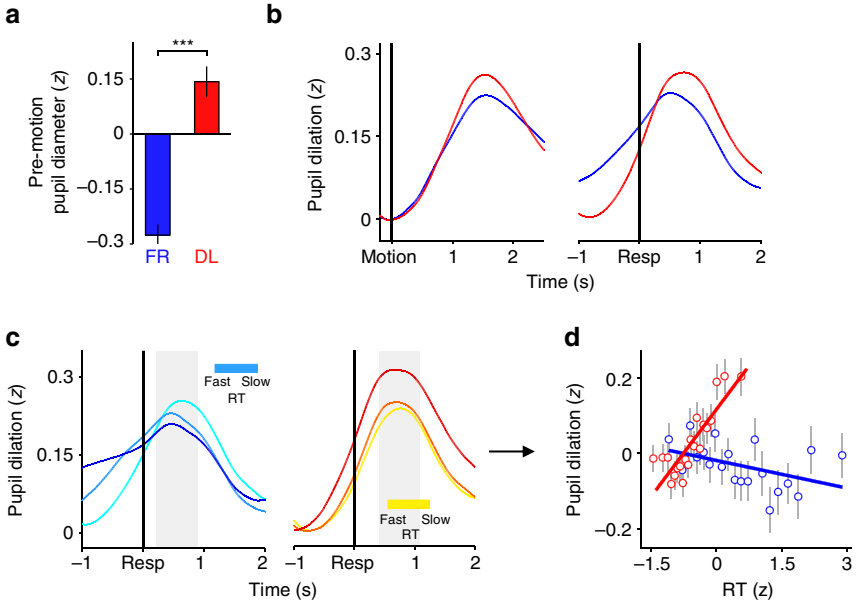

Figure 4 | Effects of deadline-induced speed pressure on pre-motion pupil diameter and evoked pupil dilation. (**a**) Effect of speed regime on unbaselined pupil diameter, as measured during the period directly preceding motion onset. Error bars = s.e.m. *** = $P<0.001$. (**b**) Evoked pupil dilation separated by speed regime and aligned to both motion onset (left) and response (right). Waveforms are relative to a pre-motion baseline. (**c**) Response-aligned pupil dilation after sorting trials by RT into 3 equal-sized bins, separately for the FR and DL conditions. (**d**) Scatterplot illustrating the linear relationships between RT and evoked pupil dilation for each speed regime. Points and error bars are mean ± s.e.m. of data that were *z*-scored within subjects, pooled across subjects and grouped into 20 bins; *z*-scoring was carried out across speed regimes to preserve main effects of speed emphasis. In **c**, shaded grey regions show measurement windows for effects shown in **d** and reported in text.

deadlines < 0.1%) and the dynamics of the evidence accumulation process. With respect to the accumulation dynamics, the activation time-series of the simulated accumulator units in the model displayed the two critical characteristics of static and time-dependent urgency that we observed under deadline in motor preparation signals in the human EEG: a baseline increase in activation during the pre-motion period (Fig. 6f); and, stronger common activation of both accumulators (reflected in a smaller difference between accumulators) at the time of commitment for slower decision times (Fig. 6g,h). These simulations suggest that global gain modulation is a plausible biophysical mechanism for generating static and time-dependent urgency in the brain.

It has recently been argued that, rather than relying on gradual evidence accumulation, decisions are determined by a more instantaneous estimate of the current sensory evidence combined with a growing urgency signal[19,23,28,52]. In our simple network model, such a regime can be approximated by constraining the effective time constant of accumulation (τ) to be particularly short (Supplementary Methods). When we enforced this constraint, the model still provided a reasonable account of behaviour and accumulation dynamics (Fig. 6d,e,h, thin grey lines). Thus, evidence accumulation with a long time constant does not appear to be a necessary prerequisite for generating the data observed presently.

Time-dependent urgency under mild speed pressure. The signatures of urgency that we report above were observed in task contexts of high speed pressure. In a final set of analyses, we examined whether the same mechanism might also be invoked in situations where speed pressure is less severe. We re-analysed data from two experiments in which subjects again made motion discrimination decisions, but without any manipulation of speed emphasis. Instead, they performed under a deadline of 1.5 s at all times and there was no explicit penalty for missed deadlines. This task feature has been employed previously in studies of human perceptual decision-making that were not designed to interrogate the mechanistic basis of SAT regulation (for example, refs 34,53).

In the first of our re-analysed studies[34] (Fig. 7a), subjects missed a low proportion of deadlines (median = 0.50 ± 0.16%), and their CAFs arrived at a mean accuracy level at the time of the deadline that was not different from chance across subjects (48.0 ± 3.9%; $t_{25} = -0.5$, $P = 0.6$). In the second study (unpublished; Fig. 7b), subjects performed under two difficulty levels and again missed very few deadlines (easy = 0.15 ± 0.13%; hard = 0.63 ± 0.15%). Moreover, despite the CAFs for each difficulty level being significantly different for almost the entire range of RTs, they converged to approximately chance accuracy at the deadline (easy: 51.9 ± 4.5%, $t_{20} = 0.4$, $P = 0.7$; hard: 46.2 ± 2.0%, $t_{20} = -1.9$, $P = 0.07$; paired-samples *t*-test for easy versus hard: $t_{20} = 1.3$, $P = 0.2$). As described previously, this repeatedly observed combination of few missed deadlines, strongly negative CAF slopes and chance performance around the time of the deadline is a hallmark of a time-dependent decision policy.

Discussion

In models of decision-making, a common assumption is that the accuracy and timing of decision commitment are determined by a context-dependent but time-invariant criterion on accumulated evidence[7–10,16]. Theoretical considerations suggest that such a time-invariant policy is sub-optimal if the potential cost of continued evidence accumulation grows with elapsed decision time, as is often the case in decision-making contexts that place a premium on fast responding[18,20]. Yet, in support of the principle of time-invariance, recent reports have suggested that human decision-makers may fail to implement a dynamic, time-variant commitment policy that would yield higher reward rates in such settings[15,29,30]. In the present study, we describe strong, convergent evidence to the contrary. Through analysis of observed behaviour, computational modelling and scalp

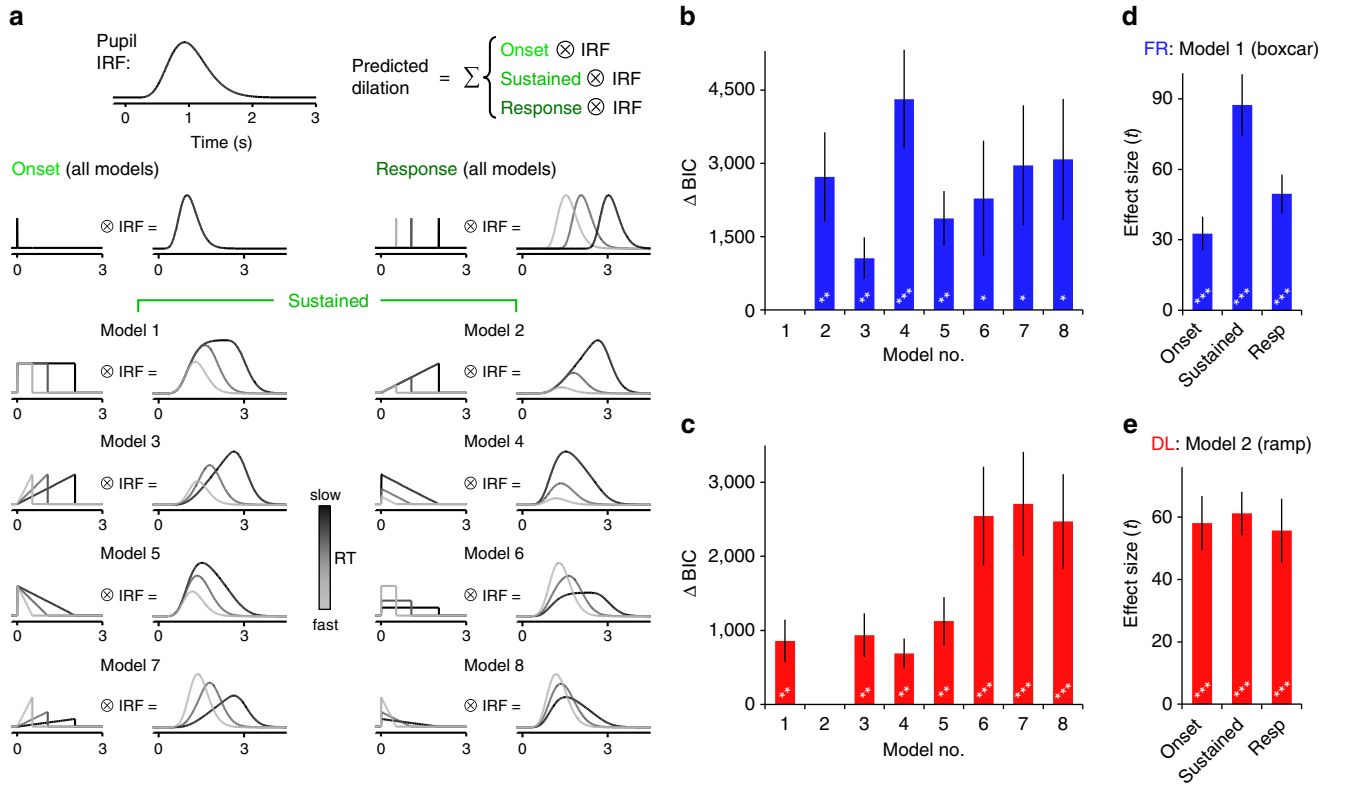

Figure 5 | Identifying the shape of the neural input to the pupil diameter system during decision formation. (**a**) Schematic depicting the analysis approach employed to parse the contributions of distinct temporal components to the observed pupil dilation waveforms and identify the likely shape of the input to the pupil system during decision formation. Three temporal components (onset, sustained, response) were convolved with a canonical pupil impulse response function (top left) and regressed onto the observed pupil diameter time series. The shape of the sustained decisional component differed across 8 candidate models and their goodness-of-fits were quantitatively compared. (**b**,**c**) BIC scores for each of the 8 models fit to the FR (**b**) and DL (**c**) data, relative to the winning model in each case (model 1 for FR; model 2 for DL). (**d**,**e**) Effect sizes of parameter estimates for each of the temporal components from the winning FR (**d**) and DL (**e**) models. Error bars = s.e.m. *** = $P<0.001$, ** = $P<0.01$, * = $P<0.05$.

electrophysiological and pupillometric findings, we show that human subjects are capable of adapting to deadline-induced speed pressure via a combination of static and time-dependent changes to their criterion on accumulated evidence.

Recent studies that applied quantitative model comparison techniques to multiple behavioural datasets provided some support for the presence of a time-dependent influence on the decision process of highly-trained monkeys, but little evidence for time-dependency in mostly naïve human subjects[15,31]. By contrast, we observed clear support for model variants with strong time-dependency in humans that, aside from brief initial training sessions, had no prior experience with the imperative task. How might this discrepancy in findings be explained? One likely contributing factor is differences in the nature of the speed pressure created by the various decision-making contexts in question. In our task, the heavy punishments levied for missed deadlines created strong, time-sensitive speed pressure that was likely sufficient to mitigate the bias toward accurate over reward-maximizing behaviour that human subjects can display can display in choice RT settings[4,54]. On the other hand, this may not have been the case in previous studies that imposed only small, implicit penalties for slow responses (in the form of foregone rewards; for example, ref. 30), or did not provide performance-related incentives at all (for example, ref. 31).

It is also possible that mild time-dependency was present in previous investigations but not identifiable in model fits to behaviour. Specifically, popular time-invariant sequential sampling models can include variability parameters that produce similar behavioural effects as moderate time-dependent changes in the decision policy[7,33,34], potentially rendering the two indistinguishable via model comparison alone. In our case, targeted analysis of overt behaviour, measured in contexts of both strong and mild deadline-induced speed pressure, revealed signatures of time-dependency that cannot, in principle, be produced solely by variability parameters. These behavioural patterns are driven by a small percentage of trials with RTs close to the deadline and in many cases may exert a negligible influence on likelihood estimates commonly used for model fitting, but can nonetheless be highly informative when attempting to arbitrate between competing mechanistic accounts. Thus, future investigations of time-dependency in the decision process might benefit from invoking a combination of formal model comparison and assessment of such behavioural trends.

A third possibility is that the duration of the deadline that we imposed, which is long relative to the sub-second deadlines in some previous studies (for example, ref. 30), was particularly well-suited to revealing signatures of time-dependency. This prospect may point to a dependence of precisely-timed within-trial adjustments of decision policy on neural systems dedicated to the estimation of relatively long temporal intervals[55] or, perhaps complementarily, to constraints on the timescale over which the neural mechanisms responsible for these time-dependent adjustments operate. We note, however, that time-dependency operating over much faster timescales has previously been reported in the animal literature[21,22].

Although our behavioural and modelling results provide strong support for the existence of adaptive, time-dependent adjustments in subjects' decision policies under deadline, the

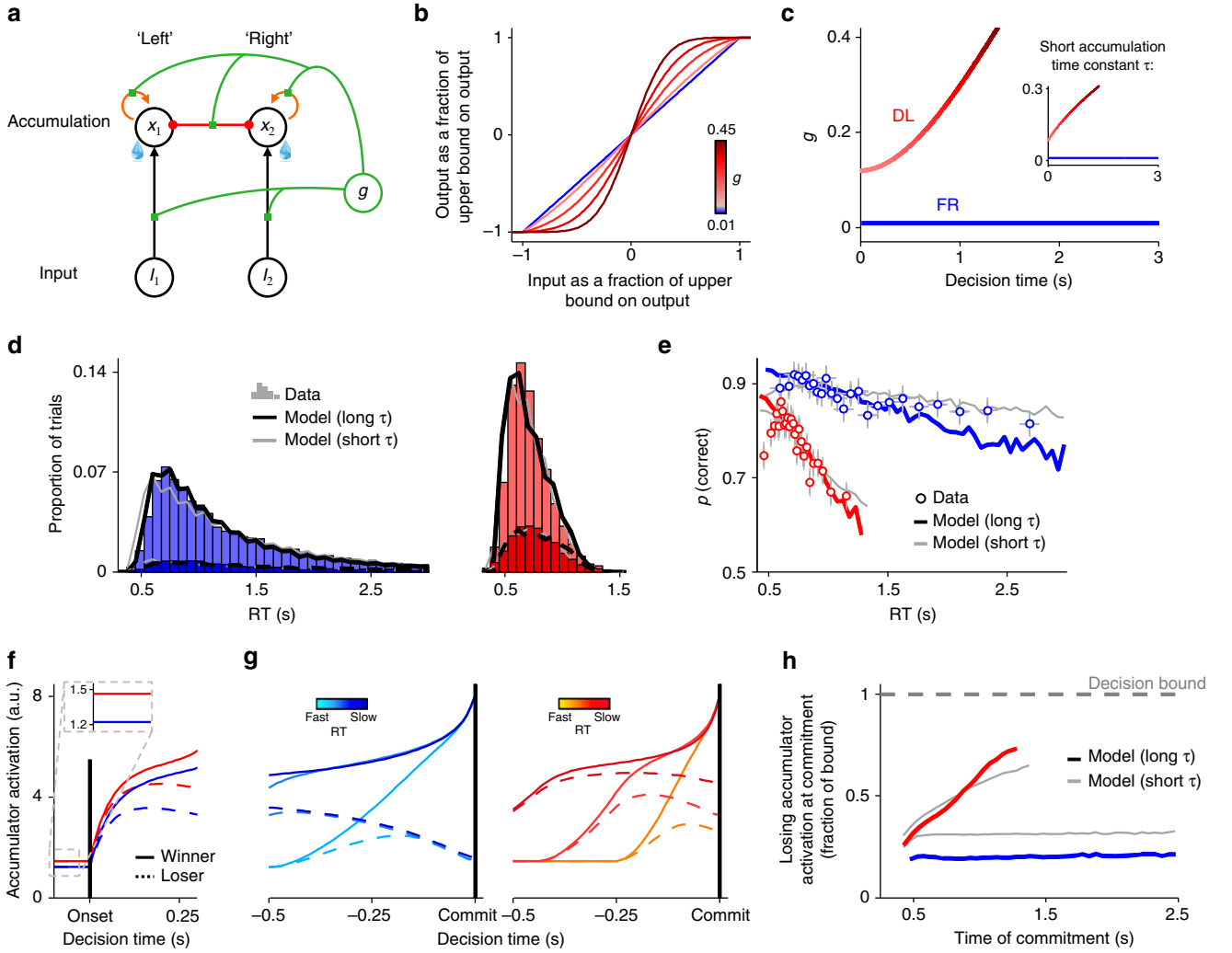

Figure 6 | A simple network model generates signatures of speed pressure via global gain modulation alone. (**a**) The model represented as a simple two-layer network in which choice-determining accumulation units are subject to recurrent excitation (orange), lateral inhibition (red) and leakage, and all connection strengths are modulated by global network gain (green). (**b**) The non-linear transfer function relating a unit's input to its corresponding output. The global gain parameter g determines the slope of this function. (**c**) g as a function of elapsed decision time in each speed regime. Informed by the pupil diameter results, there is a static offset in g at the beginning of a DL trial coupled with a time-dependent increase in the DL but not FR regime. All other model parameters are fixed across regimes. Main plot shows g time courses when the effective time constant of accumulation τ is unconstrained and takes a long value (555 ms); inset shows evolution of g when τ is constrained to equal 167 ms (see ref. 52). (**d**,**e**) Observed and fitted RT distributions and conditional accuracy functions. Points in **e** depict mean ± s.e.m. accuracy of trials sorted by RT into 25 equal-sized bins. (**f**,**g**) Activation time-courses of the winning and losing accumulator units, simulated using fitted model parameters with unconstrained τ and aligned to motion onset (**f**) and decision commitment (**g**, sorted by RT into 3 bins). Global gain modulation qualitatively produces both a pre-motion offset in activation of both accumulators and a time-dependent increase in common activation of both accumulators under deadline. (**h**) Activation of the losing accumulator at decision commitment plotted as a function of commitment time for each speed regime.

time-course of these adjustments was not strictly optimal. If afforded high flexibility of form, the optimal, reward-maximizing policy given our task design is to adopt a predominantly static criterion on accumulated evidence that steeply declines to zero at a latency determined by the subject's level of deadline timing uncertainty[20,30]. In our data, however, the shape of the observed time-dependency approximated the reward-maximizing case only if the criterion change in these calculations was constrained to be linear in time. This approximately linear trajectory was strikingly preserved across all subjects, and is similar in form to the time-dependent policy adjustments that have been observed in brain and behaviour in non-human primates[21–23]. Collectively, these findings could point to basic limitations of the neural mechanisms responsible for generating time-dependency in the decision process and, consequently, to constraints on the application of such policy adjustments for reward rate maximization in different settings.

Using lateralized 8–14 Hz oscillations in the EEG as a proxy for decision-related motor preparation[36–38], it was possible to establish that speed emphasis appeared to affect the dynamics of decision formation via a combination of static and time-dependent urgency, rather than a change in the level of the decision bound. Specifically, while the motor signals reflecting preparation for the ultimately chosen alternative reached a stereotyped pre-response level across speed regimes, we observed a deadline-induced bi-lateral increase in baseline preparation prior to decision onset, coupled with peri-decisional common activation of both effectors that was greater for slower responses. These effects have clear analogues in previous reports. In humans, functional MRI studies indicate that speed emphasis is at least

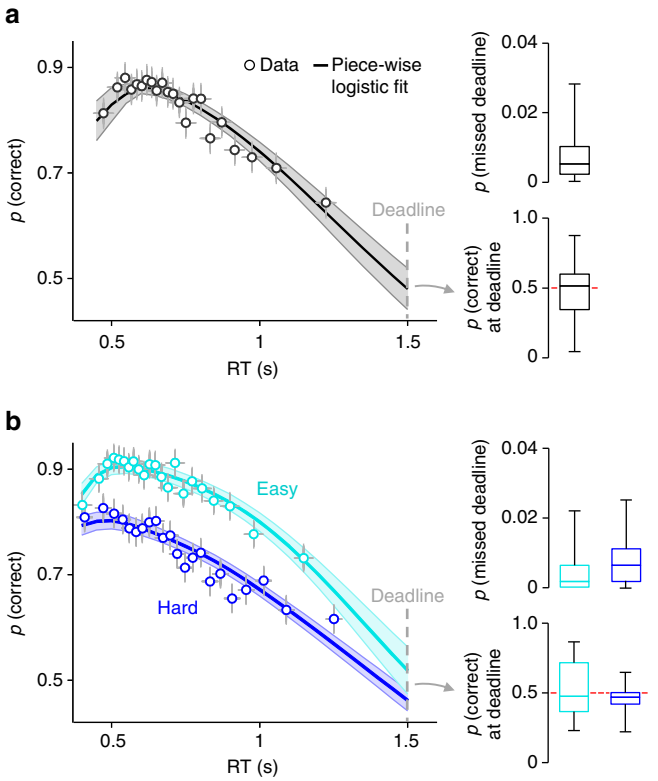

Figure 7 | Behavioural signatures of time-dependency under mild speed pressure. (**a**) Conditional accuracy function, proportion of missed deadlines and estimated accuracy at deadline in a context of mild deadline-induced speed pressure. Subjects performed the random dot motion task at a single level of discrimination difficulty. For the conditional accuracy functions, points indicate mean accuracy of trials sorted by RT into 22 equal-sized bins and line shows the mean of the best fits of piece-wise logistic regressions to each subject's single-trial data. Error bars and shaded areas indicate ± s.e.m. of data points and regression lines, respectively. Box plots show the sample median (centre line), interquartile range (box) and full range (whiskers). (**b**) Equivalent plots from a second study in which subjects performed under two levels of discrimination difficulty.

partly generated by an increase in the baseline activation of a network of decision-related brain regions (reviewed in ref. 1). In monkeys, speed pressure has similarly been shown to manifest in higher baseline firing rates of single neurons that reflect the developing decision process, but also in a time-dependent, evidence-independent increase in firing rates indicative of a growing urgency signal[21–24]. In light of the latter, our findings suggest that the mechanistic basis of SAT adjustment may be conserved across species. They also highlight that correlates of this mechanism in action are observable at the level of scalp electrophysiology.

How such urgency might be generated in the brain has been the subject of recent interest. Biophysically detailed computational analyses indicate that modulation of the gain of neural processing is one plausible mechanism for generating urgency in decision circuits[19,26,27,39,40]. Building on associations between pupil diameter and the activity of brainstem neuromodulatory systems (the LC–NA system in particular[41–43]) and the established role of these systems in global gain modulation[44–47], we provided empirical support for these ideas. We observed that pupil diameter during the pre-motion period was reliably larger under deadline and that decision-related pupil dilation increased with elapsed time specifically in this condition, thereby identifying pupillometric counterparts to the static and dynamic signatures of urgency that were observed in brain and behaviour.

Pupil-linked neuromodulatory nuclei like the LC project to almost the entire cerebral cortex and their associated neuromodulator release exerts a multiplicative influence on neural dynamics that interacts with the strength and location of ongoing processing[44,46,56]. The implication of such a general, global mechanism for urgency generation is appealing in part because it affords a simple yet very powerful means for affecting decision-making that is not specific to any one sensory input modality or effector. Indeed, global gain modulation might plausibly account for recent observations that urgency manifests not only in the firing rates of neurons that track the evolving decision process, but also in the gain of sensory inputs to decision circuits[57], in more downstream neurons involved directly in movement execution[23], and in the 'vigour' of task-irrelevant saccades during manual reaching decisions[28]. Similarly, such a global mechanism for urgency generation affords a parsimonious explanation for two potentially related observations in our data that can be viewed as distinct from effects of speed emphasis on the evidence accumulation process *per se*: the marginally quicker non-decision times under deadline; and, the deadline-induced desynchronization of baseline 8–14 Hz EEG power over occipital scalp, a phenomenon which itself has been associated with increased gain of responses in visual cortex[58].

We adapted a simple neural network model[12] to verify that global gain modulation alone can produce the behavioural and neural effects of deadline-induced speed pressure. To this end, a key qualitative effect that required reproduction was the time-dependent increase in the common activation of both accumulators under speed pressure, which we observed in EEG motor preparation signals and has also been reported in the firing rates of single neurons involved in decision formation[22,23]. Interestingly, this effect breaks the winner-take-all attractor dynamics characteristic of typical configurations of biophysically detailed spiking network models of decision-making[59,60], and in general cannot be generated via gain modulation alone in simpler models (like the DDM) that do not incorporate a recurrent excitation component. In our model, the effect was reproduced by constraining the recurrent excitation of accumulators to be stronger than the lateral inhibition between them, such that increasing network gain over time effectively heightened the dominance of excitation over inhibition and led to building activation in both accumulators. In principle, though, such an effect would also be produced by a network with more balanced excitation/inhibition in combination with stronger gain modulation for excitatory than inhibitory connections (cf. refs 39,40). Physiological data exploring potential differences in the neuromodulation of NMDA and GABAergic receptors could be highly informative about whether such dynamic changes in the ratio of excitation to inhibition occur in decision circuits.

Finally, although our analyses indicate that global gain modulation is sufficient to produce the qualitative effects of deadline-induced speed pressure, this does not preclude the existence of other sources of urgency in the brain. In particular, it has been suggested that enhanced speed pressure leads to the recruitment of cortico-basal ganglia pathways that in turn generate an effective additive input, via release from inhibition, to decision and motor circuits[1,61,62]. Such an influence could act in tandem with multiplicative gain modulation to amplify both the static and time-dependent effects of speed emphasis observed here.

Methods

Subjects. We report data from four independent cohorts of subjects. All subjects were over the age of 18, had normal or corrected-to-normal vision, and no history

of psychiatric illness or head injury. They provided written informed consent and all procedures were approved by the ethics committee of the Leiden University Institute of Psychology. Subjects received either course credit or a fixed or performance-dependent gratuity for their participation. Sample sizes were pre-planned and consistent with other studies of human decision-making, from our lab and others, that interrogated similar physiological signals and invoked similar analytical methods. Cohort-specific information is given in the Supplementary Methods.

General task procedures. All reported studies employed variants of the random dot motion (RDM) paradigm[32]. Here we report general task procedures; additional study-specific information is provided in the Supplementary Methods.

Stimuli were presented using the Psychophysics Toolbox[63] for Matlab. Subjects maintained fixation on a centrally presented cross and decided whether the dominant direction of motion of a cloud of moving dots centred on fixation was either leftward or rightward. The difficulty of these discriminations was determined by the coherence c' of the cloud of dots. Subjects indicated their decision by pressing one of two spatially compatible response keys with their left/right index fingers and were typically given feedback about the accuracy of their response on each trial. The interval between a subject's response and subsequent trial onset was randomly drawn from uniform distributions with study-specific bounds, and as such the response-to-stimulus interval did not depend on RT from the previous trial. Subjects completed initial practice and difficulty calibration routines prior to main testing, and in most cases were monetarily rewarded and punished in a performance-dependent manner under study-specific incentive schemes. Of particular note, subjects in studies 1 and 2 received 0.5¢ for every correct decision, lost 0.5¢ for every incorrect decision and, in the DL regime of these studies, lost 5¢ if they failed to respond within a temporal deadline of 1.4 s following motion onset. This relatively heavy punishment for missed deadlines serves to heighten the deadline-induced speed pressure and was implemented to mitigate the 'accuracy bias'—a prioritization of accurate decisions even at the cost of decreased reward rate—that human subjects sometimes display on choice RT tasks[4,54] and could attenuate time-dependent adjustments to decision policy[30].

In all studies, subjects attended a single testing session and discrimination difficulties were calibrated to yield approximately similar response accuracies across individuals. In study 1, 21 subjects performed 8 blocks of 180 trials at a fixed discrimination difficulty per individual (equating to 75% accuracy under deadline), with 4 blocks under a deadline of 1.4 s and 4 under free response. In study 2, 23 subjects performed 10 blocks of 90 trials, split into the same DL/FR conditions at the same subject-specific difficulty setting. In study 3, 26 subjects performed 5 blocks of 100 trials, this time all at a lower discrimination difficulty (equating to 85% accuracy) and a response deadline of 1.5 s. In study 4, 21 subjects performed 8 blocks of 160 trials, again under a 1.5 s deadline but now two difficulty levels (corresponding to 70 and 85% accuracies) that were interleaved in random order across trials within each block. In all cases subjects were familiarized with the task and encouraged to form stable estimates of the precise timing of the response deadline during practice routines.

Several task design features were implemented to minimize contamination of EEG (study 1) and pupillometric (study 2) signals: Upon response execution, coherent dot motion transitioned to purely random motion for a fixed time to avoid sensory or feedback-related transients at the time of response execution and minimize post-decisional evidence accumulation[64]; a mask of static dots was displayed during the inter-motion interval to avoid luminance-related transients at motion onset; and, during pupillometry, the inter-trial interval was extended to negate contamination of the baseline period by the previous trial's dilation response, and post-response feedback was not provided so that the dilation response was not contaminated by feedback-related processes.

All statistical tests were two-tailed. In cases where data were non-normal (as determined by the Kolmogorov–Smirnov test), non-parametric tests were used as described below.

Analysing empirical conditional accuracy functions. Single-trial logistic regression was used to estimate mean accuracy as a function of RT (the CAF), for each task condition and subject. To account for both the dominant decreasing portion of the observed CAFs and an initial increasing portion due to a small percentage of inaccurate premature responses, we constructed an algorithm that minimizes the combined sum of squared errors of piece-wise logistic regressions of accuracy (1 = correct, 0 = error) onto RT, splitting trials before and after a temporal inflection point α such that

$$P_{\text{correct}} = \begin{cases} \left(1+e^{-(\beta_0+\beta_1\times(\text{RT}-\alpha))}\right)^{-1}, & \text{RT}-\alpha \le 0 \\ \left(1+e^{-(\beta_0+\beta_2\times(\text{RT}-\alpha))}\right)^{-1}, & \text{RT}-\alpha > 0 \end{cases} \quad (1)$$

Here β_0 is accuracy at α, β_1 is the slope of the CAF before α, and β_2 is the slope of the CAF after α. β_1 was constrained to be ≥ 0 to reflect the fact that the left segment of the piece-wise fit should only account for the initial increasing portion of the CAF. This model was fit using Nelder-Mead simplex minimization to estimate the β_0, β_1 and β_2 parameters while conducting an exhaustive search of possible α values (step-size = 10 ms ending at 1 s). Whichever piece-wise segment is fit first determines β_0 and thus constrains the fit of the remaining segment; therefore, the algorithm was run twice (left segment fit first and right segment fit first) for each α to find the true minimum[30]. In fits to FR trials from task 1, all RTs longer than 5 s were excluded.

To estimate accuracy at the time of the deadline (Fig. 1c), we used the piece-wise regression fits for each subject to calculate accuracy when RT = 1.4 s. The low number of trials immediately preceding the deadline prohibits a precise characterization of the shape of the empirical CAF at this time point. However, the above single-trial regression approach allows for a reasonable approximation by exploiting consistencies in the temporal evolution of the CAF. The appropriateness of this approach relies on any change in decision policy being gradual rather than abrupt, which appeared to be the case in our data given the smooth right tails of the DL RT distributions (Fig. 1b) and the shapes of the fitted urgency signals (Fig. 3b).

EEG acquisition and analysis. Continuous EEG was acquired from the first study cohort using an ActiveTwo system (BioSemi, The Netherlands) from 64 scalp electrodes, configured to the standard 10/20 setup and digitized at 512 Hz. Eye movements were recorded using two electrodes positioned above and below the left eye and two electrodes positioned at the outer canthus of each eye. EEG data were processed in Matlab via custom scripting and subroutines from the EEGLAB toolbox[65]. We describe the full EEG preprocessing pipeline in Supplementary Methods. In brief, we used Morlet wavelet convolution to estimate the power of effector-specific μ (8–14 Hz) oscillations, which we then employed as an index of decision-related motor preparation[36–38]. Pre-stimulus μ power was measured as the mean power from −0.3 to −0.1 s preceding motion onset. 8–14 Hz power in the human EEG is subject to a prominent decrease in the immediate post-stimulus period that is generated over lateral occipital scalp (Fig. 2b, middle inset) but spreads anteriorly and contaminates early portions of the motor preparation signals of interest here. For this reason, we restricted our analyses of post-onset μ signals to the period immediately preceding response execution, which is less susceptible to contamination by this early occipital response. Pre-response μ was measured as the mean power from −0.17 to −0.05 s preceding response execution (a window that was centred on the latency of peak desynchronization in the response-aligned grand-averages and chosen in a manner that was orthogonal to potential RT and condition × RT effects; Fig. 2b, right).

We interrogated relationships between μ power and decision-making behaviour via a series of single-trial within-subjects regression models that are described in the Results section and specified in full in Supplementary Methods. For all analyses, FR trials with RT > 5 s were not included. Additionally, to mitigate the influence of the stereotyped stimulus-evoked occipital response (Fig. 2b, middle inset) on the motor μ signals of interest here and also exclude 'fast guesses' from analysis, we discarded trials with RT < 0.5 s from all EEG analyses. For all models, the group-level significance of effects represented by individual regression coefficients (β_i) was tested via one-sample t-test (H_0: $\beta_i = 0$).

In one analysis, we examined μ signals for evidence of a time-dependent influence on motor preparation that varied with speed pressure. With the decision bound fixed, the difference in activation between accumulators at the time of decision commitment can provide a proxy for the strength of an additive urgency signal. Specifically, if the decision process is driven by evidence accumulation without urgency in a winner-take-all competitive network (for example, ref. 59), then the winning accumulator will inhibit the losing accumulator and the difference in their activations will be large by the time the decision bound is reached. On the other hand, if both accumulators also receive additional, evidence-independent input due to urgency, then the common activation of both accumulators at the time of commitment should increase in proportion to the strength of the urgency signal at that time and there will be less of a difference between accumulators when urgency is stronger[22,23]. Thus, the shape of the urgency signal over time can be approximated by examining, across all levels of RT, either the raw amplitude of the losing accumulator at the time of commitment, or the difference in activation between accumulators at that time. We focus on the latter because a difference metric is, in principle, more robust to any RT-dependent contamination of μ signals by the strong bi-lateral occipital response described above (Fig. 2b, middle inset). However, we also examined the pre-response amplitude of the ipsi-lateral μ signal alone for time-dependency, and this analysis yielded similar effects (Supplementary Fig. 6).

We also explored the relationship between decision-making behaviour and effector-specific power in the β frequency band (14–30 Hz), which has also been linked to decision-making[36–38]. However, β power did not exhibit several of the critical effects that we identified in the μ signals (Supplementary Fig. 7).

Pupillometric acquisition and analysis. Pupil diameter and gaze position of the second study cohort were recorded at a sampling rate of 250 Hz using an Eyelink 1000 eye-tracker (SR Research, Canada), and analysed in Matlab. After data cleaning and artifact rejection (see Supplementary Methods), we employed a paired-samples t-test and single-trial within-subjects regressions to examine speed regime effects on pre-motion pupil diameter and post-onset pupil dilation, respectively. Pre-motion pupil diameter was measured as the mean, unbaselined pupil diameter from −0.2 to 0 s relative to motion onset. Evoked pupil dilation was measured as the mean pupil diameter within a 0.7 s window centred on the latency of peak dilation in the response-aligned grand-average waveforms from each speed regime (0.55 s post-response in FR, 0.75 s post-response in DL; Fig. 4b),

baselined relative to the pre-stimulus interval, though we also show that the reported effects are robust to different measurement windows (Supplementary Fig. 4).

The phasic input to the peripheral system controlling pupil diameter was modelled as a linear combination of three temporal components: transients at motion onset and response, and a sustained component throughout the intervening period[49,50]. For each subject and speed emphasis condition, eight different models were constructed in which the sustained component took one of the following shapes (Fig. 5a): (1) a boxcar with constant amplitude throughout the decision interval; (2) a linear up-ramp that grew in amplitude with increasing decision time; (3) a ramp-to-threshold; (4) a linear decay with a starting amplitude that was larger for slower RTs but whose amplitude always terminated at zero; (5) a linear decay-to-threshold which began at a fixed amplitude and terminated at zero; and, (6–8) versions of the boxcar, up-ramp and down-ramp in which the sustained component for each trial was normalized by the number of samples in that trial's decision interval, thereby negatively modulating these components by RT. To fit each model, a vector of concatenated pupil dilation waveforms, from 0.2 s pre-stimulus to 2.5 s post-response, was regressed onto a general linear model composed of the three temporal components (onset, sustained, response) convolved with a pupil impulse response function[51]:

$$h(t) = t^{w} \times e^{-t(w/t_{max})} \quad (2)$$

where $w = 10.1$ and $t_{max} = 930$ ms (matching the function used in refs 49–51, though the key findings were robust to specific parameter combinations; Supplementary Fig. 5). Model fit was assessed using the Bayes Information Criterion (BIC) for models estimated via least squares:

$$BIC = n + n\log(2\pi) + \log(SSR/n) + (k+1)\log(n) \quad (3)$$

where n is the number of samples, SSR is the residual sum of squares, and k is the number of free parameters. The relative goodness of fit between two given models was assessed non-parametrically by subjecting difference values (BIC_1-BIC_2) to Wilcoxon signed rank tests.

Drift diffusion modelling. Behavioural data from the study 1 cohort were fit with several versions of the DDM for two-alternative decisions[10], both with and without an urgency component. In its most basic form, the DDM assumes that noisy sensory evidence is accumulated from a starting point z at drift rate v and a decision is made when a criterial amount of cumulative evidence reaches one of two opposing boundaries corresponding to either choice option. The distance between boundaries is the boundary separation a, while the model ascribes all non-decision-related processing to a non-decision time parameter t_{er}. Noise in the evidence is determined by s, the s.d. of a zero-mean Gaussian distribution, and is fixed at 0.1 to scale all other parameters[66]. Given any combination of the above parameters, the DDM yields a flat CAF and thus cannot account for the negative CAF slopes that we observed in our data. However, including between-trial variability in drift rate (normally distributed with s.d. $= \eta$) allows the model to produce decreasing CAFs[33,34], and so we also included this parameter in our model fits. In all models, z was fixed at $a/2$. Thus, what we refer to as the 'standard DDM' had, at a minimum, four free parameters (v, η, a, t_{er}).

Informed by our EEG findings, we also considered DDM variants that incorporate an additive urgency component. In the 'urgency DDM', decisions are determined by the states of two perfectly anti-correlated accumulators that are subject to regular drift diffusion, and are each summed with the same time-varying, evidence-independent quantity (the urgency signal). A decision is made when the total activation (diffusion + urgency) of one of the accumulators passes a common decision bound (fixed at 1 for all conditions and subjects). The shape of the urgency signal was parameterized by a logistic function:

$$u(t) = u_0 + \left(1 - e^{-(t/\lambda)^k}\right) \quad (4)$$

where $u(t)$ is the magnitude of the urgency at decision time t, u_0 is the static component of the urgency (that is, the value of u when $t = 0$), and k and λ are shape and scale parameters that determine the shape of the time-dependent component of the urgency. The logistic function was chosen because it can produce a variety of different shapes of urgency signal (concave, convex, approximately linear, flat) using few free parameters. Although conceptually distinct, this urgency model is mathematically identical to a model in which the standard DDM is coupled with time-varying decision bounds. The urgency DDM had a minimum of five free parameters (v, t_{er}, u_0, k, λ), or six in cases where η was also included.

We fit a number of models with varying parameter constraints (Supplementary Table 1) and estimated parameters for each model and subject using maximum likelihood estimation procedures that are described in the Supplementary Methods. Of particular note, for the urgency DDM we invoked a method for analytically deriving first passage time densities through continuously differentiable time-varying bounds[67]. This approach is based on the analysis of renewal equations and described in detail by Smith[68] and Zhang *et al.*[69]. The Supplementary Methods also contain a detailed description of our approach for estimating the optimal, reward-maximizing time-dependent urgency signals, given our task, for a representative set of time-invariant parameters.

Leaky competing accumulator modelling. To interrogate effects of global gain modulation on decision-making, a modelling approach must be employed that allows basic features of neural information processing, such as the relative strength of recurrent excitation and lateral inhibition, to be dissociated; these properties of a neural network, which are not distinguished in the more abstract DDM, determine the nature of effects of gain modulation on accumulation dynamics and decision-making behaviour[39,40]. We therefore modelled gain modulation by adapting the LCA model[12], which is built upon such principles of neural computation and offers a tractable means of interrogating gain effects without the level of complexity inherent in more biophysically detailed models of decision-related neural population dynamics[39,40,59]. Note that we did not employ this model for earlier quantitative model comparison because, despite its simplicity relative to more biophysically plausible neural networks, it is under-constrained (see Supplementary Methods).

In the two-alternative LCA model, decision-making is driven by a simple two-layer neural network consisting of two units over which external input is represented, and two accumulator units, one for each response alternative, that determine choice (Fig. 6a). Each unit, which represents a population of functionally equivalent neurons, is characterized by two variables: its activation, which captures the net input to the unit, and its output, which is related to activation via a nonlinear transfer function (see below). The activation values of the first (correct) and second (incorrect) input units are I_1 and I_2, respectively, and their associated outputs to the accumulator units are $f(I_1)$ and $f(I_2)$. The momentary change in the activation of each accumulator unit x_i can be approximated by the following finite difference equations[12]:

$$\begin{aligned} \Delta x_1 &= f(I_1) - \lambda x_1 + \alpha f(x_1) - \beta f(x_2) + f(N(0,\sigma)) \\ \Delta x_2 &= f(I_2) - \lambda x_2 + \alpha f(x_2) - \beta f(x_1) + f(N(0,\sigma)) \end{aligned} \quad (5)$$

and the accumulator units are subject to a lower bound on activation such that:

$$\begin{aligned} x_1(t+1) &= \max(0, x_1(t) + \Delta x_1) \\ x_2(t+1) &= \max(0, x_2(t) + \Delta x_2) \end{aligned} \quad (6)$$

In Equation (5), λ represents the leak or decay of activation over time, α represents recurrent excitation, β represents lateral inhibition, and $N(0,\sigma)$ is a zero-mean Gaussian-distributed noise term with s.d. $= \sigma$. A decision is made in the model when the activation of one of the accumulator units exceeds a decision bound A.

Note that, with the exception of the leak, every term that contributes to Δx_i in Equation (5) is passed through the transfer function relating a unit's activation to its output. Varying the slope of this function provides a natural way to implement global gain modulation in the LCA. In accordance with extensive previous modelling work (for example, refs 27,45,47,70,71), we assumed that the transfer function is sigmoidal in shape. The sigmoid places upper and lower bounds on output and thus prevents runaway activation in cases where the effective recurrent excitation is greater than the leakage (i.e. $\alpha f(x_i) - \beta f(x_{i' \neq i}) > \lambda x_i$), which can happen when gain is high. We favored a transfer function that becomes linear within the range of possible outputs as gain approaches 0, thus approximating the threshold-linear function employed in the original LCA model[12], and step-like as gain approaches ∞. This function took the following form[70]:

$$f\left(x \mid 0, \frac{1}{g}\right) = \begin{cases} -\theta, & x \leq -\theta \\ -\theta + 2\theta \dfrac{\int_{-\theta}^{x} \varphi\left(y \mid 0, \frac{1}{g}\right) dy}{\int_{-\theta}^{\theta} \varphi\left(y \mid 0, \frac{1}{g}\right) dy}, & -\theta \leq x \leq \theta \\ \theta, & x > \theta \end{cases} \quad (7)$$

where $\varphi(0,1/g)$ is the cumulative function of a normal distribution with mean $= 0$ and s.d. $= 1/g$, and θ determines the symmetric upper and lower bounds on output. The gain parameter g determines the steepness of the non-linearity in the function (Fig. 6b).

Informed by our pupillometric findings, we realized urgency through global gain modulation in this adapted LCA model by allowing both the baseline offset and within-trial time-varying trajectory of the g parameter to vary with speed regime (Fig. 6c). All other model parameters were fixed across speed regimes. Full specifications of the remaining model parameters, fitting procedures, and approach used for simulating accumulator time-series are all provided in the Supplementary Methods.

Data availability. The data and computer code that support the findings of this study are available from the corresponding author on request.

References

1. Bogacz, R., Wagenmakers, E. J., Forstmann, B. U. & Nieuwenhuis, S. The neural basis of the speed-accuracy tradeoff. *Trends Neurosci.* **33,** 10–16 (2010).
2. Heitz, R. P. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Front Neurosci.* **8,** 150 (2014).
3. Wickelgren, W. A. Speed-accuracy tradeoff and information processing dynamics. *Acta Psychol. (Amst)* **41,** 67–85 (1977).
4. Balci, F. *et al.* Acquisition of decision making criteria: reward rate ultimately beats accuracy. *Atten. Percept. Psychophys.* **73,** 640–657 (2011).

5. Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113,** 700–765 (2006).
6. Gold, J. I. & Shadlen, M. N. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36,** 299–308 (2002).
7. Brown, S. D. & Heathcote, A. The simplest complete model of choice response time: linear ballistic accumulation. *Cognit. Psychol.* **57,** 153–178 (2008).
8. Forstmann, B. U., Ratcliff, R. & Wagenmakers, E. J. Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annu. Rev. Psychol.* **67,** 641–666 (2016).
9. Palmer, J., Huk, A. C. & Shadlen, M. N. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vis.* **5,** 376–404 (2005).
10. Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion decision model: current issues and history. *Trends Cogn. Sci.* **20,** 260–281 (2016).
11. Shadlen, M. N. & Kiani, R. Decision making as a window on cognition. *Neuron* **80,** 791–806 (2013).
12. Usher, M. & McClelland, J. L. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* **108,** 550–592 (2001).
13. Kelly, S. P. & O'Connell, R. G. The neural processes underlying perceptual decision making in humans: recent progress and future directions. *J. Physiol. Paris* **109,** 27–37 (2015).
14. Forstmann, B. U. *et al.* Striatum and pre-SMA facilitate decision-making under time pressure. *Proc. Natl Acad. Sci. USA* **105,** 17538–17542 (2008).
15. Hawkins, G. E., Forstmann, B. U., Wagenmakers, E. J., Ratcliff, R. & Brown, S. D. Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *J. Neurosci.* **35,** 2476–2484 (2015).
16. Reddi, B. A. & Carpenter, R. H. The influence of urgency on decision time. *Nat. Neurosci.* **3,** 827–830 (2000).
17. van Maanen, L. *et al.* Neural correlates of trial-to-trial fluctuations in response caution. *J. Neurosci.* **31,** 17488–17495 (2011).
18. Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N. & Pouget, A. The cost of accumulating evidence in perceptual decision making. *J. Neurosci.* **32,** 3612–3628 (2012).
19. Thura, D., Beauregard-Racine, J., Fradet, C. W. & Cisek, P. Decision making by urgency gating: theory and experimental support. *J. Neurophysiol.* **108,** 2912–2930 (2012).
20. Frazier, P. & Yu, A. J. Sequential hypothesis testing under stochastic deadlines. *Adv. Neural Inf. Process Syst.* **20,** 465–472 (2008).
21. Churchland, A. K., Kiani, R. & Shadlen, M. N. Decision-making with multiple alternatives. *Nat. Neurosci.* **11,** 693–702 (2008).
22. Hanks, T., Kiani, R. & Shadlen, M. N. A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *eLife* **3,** e02260 (2014).
23. Thura, D. & Cisek, P. Modulation of premotor and primary motor cortical activity during volitional adjustments of speed-accuracy trade-offs. *J. Neurosci.* **36,** 938–956 (2016).
24. Heitz, R. P. & Schall, J. D. Neural mechanisms of speed-accuracy tradeoff. *Neuron* **76,** 616–628 (2012).
25. Cisek, P., Puskas, G. A. & El-Murr, S. Decisions in changing conditions: the urgency-gating model. *J. Neurosci.* **29,** 11560–11571 (2009).
26. Ditterich, J. Evidence for time-variant decision making. *Eur. J. Neurosci.* **24,** 3628–3641 (2006).
27. Standage, D., You, H., Wang, D. H. & Dorris, M. C. Gain modulation by an urgency signal controls the speed-accuracy trade-off in a network model of a cortical decision circuit. *Front. Comput. Neurosci.* **5,** 7 (2011).
28. Thura, D., Cos, I., Trung, J. & Cisek, P. Context-dependent urgency influences speed-accuracy trade-offs in decision-making and movement execution. *J. Neurosci.* **34,** 16442–16454 (2014).
29. Boehm, U., Hawkins, G. E., Brown, S., van Rijn, H. & Wagenmakers, E. J. Of monkeys and men: Impatience in perceptual decision-making. *Psychon. Bull. Rev.* **23,** 738–749 (2016).
30. Karsilar, H., Simen, P., Papadakis, S. & Balci, F. Speed accuracy trade-off under response deadlines. *Front. Neurosci.* **8,** 248 (2014).
31. Voskuilen, C., Ratcliff, R. & Smith, P. L. Comparing fixed and collapsing boundary versions of the diffusion model. *J. Math. Psychol.* **73,** 59–79 (2016).
32. Newsome, W. T., Britten, K. H. & Movshon, J. A. Neuronal correlates of a perceptual decision. *Nature* **3841,** 52–54 (1989).
33. Ratcliff, R. & Rouder, J. N. Modeling response times for two-choice decisions. *Psychol. Sci.* **9,** 347–356 (1998).
34. Murphy, P. R., Vandekerckhove, J. & Nieuwenhuis, S. Pupil-linked arousal determines variability in perceptual decision making. *PLoS Comput. Biol.* **10,** e1003854 (2014).
35. Bowman, N. E., Kording, K. P. & Gottfried, J. A. Temporal integration of olfactory perceptual evidence in human orbitofrontal cortex. *Neuron* **75,** 916–927 (2012).
36. de Lange, F. P., Rahnev, D. A., Donner, T. H. & Lau, H. Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *J. Neurosci.* **33,** 1400–1410 (2013).
37. Donner, T. H., Siegel, M., Fries, P. & Engel, A. K. Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Curr. Biol.* **19,** 1581–1585 (2009).
38. Gould, I. C., Nobre, A. C., Wyart, V. & Rushworth, M. F. Effects of decision variables and intraparietal stimulation on sensorimotor oscillatory activity in the human brain. *J. Neurosci.* **32,** 13805–13818 (2012).
39. Eckhoff, P., Wong-Lin, K. & Holmes, P. Optimality and robustness of a biophysical decision-making model under norepinephrine modulation. *J. Neurosci.* **29,** 4301–4311 (2009).
40. Niyogi, R. K. & Wong-Lin, K. Dynamic excitatory and inhibitory gain modulation can produce flexible, robust and optimal decision-making. *PLoS Comput. Biol.* **9,** e1003099 (2013).
41. Joshi, S., Li, Y., Kalwani, R. M. & Gold, J. I. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* **89,** 221–234 (2016).
42. Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H. & Balsters, J. H. Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum. Brain Mapp.* **35,** 4140–4154 (2014).
43. Varazzani, C., San-Galli, A., Gilardeau, S. & Bouret, S. Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. *J. Neurosci.* **35,** 7866–7877 (2015).
44. Aston-Jones, G. & Cohen, J. D. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28,** 403–450 (2005).
45. Eldar, E., Cohen, J. D. & Niv, Y. The effects of neural gain on attention and learning. *Nat. Neurosci.* **16,** 1146–1153 (2013).
46. Mather, M., Clewett, D., Sakaki, M. & Harley, C. W. Norepinephrine ignites local hot spots of neuronal excitation: how arousal amplifies selectivity in perception and memory. *Behav. Brain Sci.* 1–100 (2015).
47. Servan-Schreiber, D., Printz, H. & Cohen, J. D. A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science* **249,** 892–895 (1990).
48. Warren, C. M. *et al.* Catecholamine-mediated increases in gain enhance the precision of cortical representations. *J. Neurosci.* **36,** 5699–5708 (2016).
49. de Gee, J. W., Knapen, T. & Donner, T. H. Decision-related pupil dilation reflects upcoming choice and individual bias. *Proc. Natl Acad. Sci. USA* **111,** E618–E625 (2014).
50. Lempert, K. M., Chen, Y. L. & Fleming, S. M. Relating pupil dilation and metacognitive confidence during auditory decision-making. *PLoS ONE* **10,** e0126588 (2015).
51. Hoeks, B. & Levelt, W. J. M. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behav. Res. Methods Instrum. Comput.* **25,** 16–26 (1993).
52. Carland, M. A., Marcos, E., Thura, D. & Cisek, P. Evidence against perfect integration of sensory information during perceptual decision making. *J. Neurophysiol.* **115,** 915–930 (2016).
53. Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W. & Forstmann, B. U. Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J. Neurosci.* **32,** 2335–2343 (2012).
54. Simen, P. *et al.* Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *J. Exp. Psychol. Hum. Percept. Perform.* **35,** 1865–1897 (2009).
55. Lewis, P. A. & Miall, R. C. Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Curr. Opin. Neurobiol.* **13,** 250–255 (2003).
56. Berridge, C. W. & Waterhouse, B. D. The locus coeruleus-noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Res. Brain Res. Rev.* **42,** 33–84 (2003).
57. Heitz, R. P. & Schall, J. D. Neural chronometry and coherency across speed-accuracy demands reveal lack of homomorphism between computational and neural mechanisms of evidence accumulation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368,** 20130071 (2013).
58. Chaumon, M. & Busch, N. A. Prestimulus neural oscillations inhibit visual perception via modulation of response gain. *J. Cogn. Neurosci.* **26,** 2514–2529 (2014).
59. Wang, X. J. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36,** 955–968 (2002).
60. Roxin, A. & Ledberg, A. Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Comput. Biol.* **4,** e1000046 (2008).
61. Forstmann, B. U. *et al.* Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proc. Natl Acad. Sci. USA* **107,** 15916–15920 (2010).

62. Herz, D. M., Zavala, B. A., Bogacz, R. & Brown, P. Neural correlates of decision thresholds in the human subthalamic nucleus. *Curr. Biol.* **26,** 916–920 (2016).
63. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10,** 433–436 (1997).
64. Murphy, P. R., Robertson, I. H., Harty, S. & O'Connell, R. G. Neural evidence accumulation persists after choice to inform metacognitive judgments. *eLife* **4,** e11946 (2015).
65. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134,** 9–21 (2004).
66. Ratcliff, R. & Tuerlinckx, F. Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychon. Bull. Rev.* **9,** 438–481 (2002).
67. Buonocore, A., Giorno, V., Nobile, A. & Ricciardi, L. On the two-boundary first-crossing-time problem for diffusion processes. *J. Appl. Probab.* **27,** 102–114 (1990).
68. Smith, P. L. Stochastic dynamic models of response time and accuracy: a foundational primer. *J. Math. Psychol.* **44,** 408–463 (2000).
69. Zhang, S., Lee, M. D., Vandekerckhove, J., Maris, G. & Wagenmakers, E. J. Time-varying boundaries for diffusion models of decision making and response time. *Front Psychol.* **5,** 1364 (2014).
70. Cheadle, S. *et al.* Adaptive gain control during human perceptual choice. *Neuron* **81,** 1429–1441 (2014).
71. Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J. & Aston-Jones, G. The role of locus coeruleus in the regulation of cognitive performance. *Science* **283,** 549–554 (1999).

Acknowledgements

The authors thank Redmond O'Connell and Konstantinos Tsetsos for helpful discussions. This work was supported by an ERC Consolidator Grant awarded to S.N.

Author contributions

P.R.M. and S.N. conceived and designed the experiments. E.B. and P.R.M. collected the data. P.R.M. analysed the data and fit the models. P.R.M. and S.N. wrote the manuscript and all authors approved the final version.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

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

How to cite this article: Murphy, P. R. *et al.* Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nat. Commun.* **7,** 13526 doi: 10.1038/ncomms13526 (2016).

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

In-group defense, out-group aggression, and coordination failures in intergroup conflict

Carsten K. W. De Dreu[a,b,1], Jörg Gross[a,b], Zsombor Méder[c], Michael Giffin[a], Eliska Prochazkova[a], Jonathan Krikeb[b], and Simon Columbus[d]

[a]Institute of Psychology, Leiden University, 2333 AK Leiden, The Netherlands; [b]Center for Experimental Economics and Political Decision Making, University of Amsterdam, 1018 WB Amsterdam, The Netherlands; [c]Humanities, Arts and Social Sciences, Singapore University of Technology and Design, Singapore 487372; and [d]Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, 1081 BT Amsterdam, The Netherlands

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved July 13, 2016 (received for review March 29, 2016)

Intergroup conflict persists when and because individuals make costly contributions to their group's fighting capacity, but how groups organize contributions into effective collective action remains poorly understood. Here we distinguish between contributions aimed at subordinating out-groups (out-group aggression) from those aimed at defending the in-group against possible out-group aggression (in-group defense). We conducted two experiments in which three-person aggressor groups confronted three-person defender groups in a multiround contest game (n = 276; 92 aggressor–defender contests). Individuals received an endowment from which they could contribute to their group's fighting capacity. Contributions were always wasted, but when the aggressor group's fighting capacity exceeded that of the defender group, the aggressor group acquired the defender group's remaining resources (otherwise, individuals on both sides were left with the remainders of their endowment). In-group defense appeared stronger and better coordinated than out-group aggression, and defender groups survived roughly 70% of the attacks. This low success rate for aggressor groups mirrored that of group-hunting predators such as wolves and chimpanzees (n = 1,382 cases), hostile takeovers in industry (n = 1,637 cases), and interstate conflicts (n = 2,586). Furthermore, whereas peer punishment increased out-group aggression more than in-group defense without affecting success rates (Exp. 1), sequential (vs. simultaneous) decision-making increased coordination of collective action for out-group aggression, doubling the aggressor's success rate (Exp. 2). The relatively high success rate of in-group defense suggests evolutionary and cultural pressures may have favored capacities for cooperation and coordination when the group goal is to defend, rather than to expand, dominate, and exploit.

competition | parochial altruism | coordination | collective action | intergroup relations

Human history is marked by intergroup conflict. From tribal warfare in the Holocene to Viking raids in medieval times, to terrorist attacks in current times, small groups of often no more than a handful of individuals organize for collective violence and aggression. Individuals within such groups contribute, at sometimes exceedingly high personal cost, to their group's capacity to fight other groups (1–5), and in doing so, individuals and their groups waste resources and people and create imprints on collective memories that affect intergroup relations for generations to come (6–10).

Given the risk for injury and death, and the collective wastefulness of intergroup conflict, it may seem puzzling that people self-sacrifice and make costly contributions to their group's fighting capacity. However, by contributing to intergroup aggression, individuals enable their groups to subordinate rivaling out-groups and absorb their resources (3, 4), something from which individual group members benefit too. Indeed, groups that most effectively elicit contributions from their members are most likely to be victorious, and perhaps intergroup competition and conflict pressure individuals to contribute to intergroup violence (1, 3, 5, 11, 12) and its supporting institutions (8, 9, 13, 14).

That intergroup conflict elicits self-sacrificial contributions to one's group's fighting capacity has been robustly revealed in experiments using N-person (intergroup) prisoner's dilemma (4, 5, 15–17) or price-contest games (18–21). What cannot be derived from these setups, however, is whether individuals self-sacrifice to (*i*) defend their in-group against out-group aggression; (*ii*) to aggressively exploit and subordinate the out-group; or (*iii*) because of some combination of both reasons (5, 9, 10, 22, 23). In addition, it is unclear how the willingness to defend the in-group relates to the willingness to aggress out-groups. These issues are nontrivial because tendencies for in-group defense and out-group aggression are often differentially dispersed between opposing groups. From group-hunting by lions, wolves, or killer whales (24, 25), to groups of chimpanzees raiding their neighbors (11), to hostile takeovers in the marketplace (26), and to territorial conflicts within and between nation states (27), intergroup conflict is often a clash between the antagonist's out-group aggression and the opponent's in-group defense (23, 28). Second, in-group defense and out-group aggression appear to have distinct neurobiological origins (5, 29–31), and may thus recruit different within-group dynamics (4, 28). Whereas self-defense is impulsive and relies on brain structures involved in threat signaling and emotion regulation, offensive aggression is more instrumental and conditioned by executive control (29–31). Third, the motivation to avoid loss is stronger than the search for gain (32, 33), suggesting that individuals more readily contribute to defensive, rather than offensive, aggression. Finally, self-sacrifice in combat is publicly rewarded more (e.g., with a Medal of Honor) when it served in-group defense rather than out-group aggression (34). Accordingly, in-group defense may emerge more spontaneously, and individuals may be more intrinsically motivated to contribute to in-group defense than to out-group aggression.

Significance

Across a range of domains, from group-hunting predators to laboratory groups, companies, and nation states, we find that out-group aggression is less successful because it is more difficult to coordinate than in-group defense. This finding explains why appeals for defending the in-group may be more persuasive than appeals to aggress a rivaling out-group and suggests that (third) parties seeking to regulate intergroup conflict should, in addition to reducing willingness to contribute to one's group's fighting capacity, undermine arrangements for coordinating out-group aggression, such as leadership, communication, and infrastructure.

Author contributions: C.K.W.D.D., J.G., and S.C. designed research; J.G., M.G., E.P., J.K., and S.C. performed research; J.K., M.G., and E.P. collected data; Z.M. contributed new reagents/analytic tools; C.K.W.D.D., J.G., Z.M., M.G., E.P., J.K., and S.C. analyzed data; and C.K.W.D.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: c.k.w.de.dreu@fsw.leidenuniv.nl.

If in-group defense is indeed more intrinsically motivating and spontaneous, groups preparing for in-group defense should face fewer noncontributors than groups preparing for out-group aggression. Aggressor groups should thus have higher within-group dispersion in contributions and may have greater difficulty organizing adequate out-group aggression. This collective action problem in aggressor groups may emerge because of motivation failure (individuals are less willing to contribute to out-group aggression than to in-group defense), or it may be the result of poor coordination (it is more difficult to coordinate and align individual contributions to effectively aggress a rivaling group than it is to raise a proper in-group defense).

We examined these possibilities and their consequences for conflict trajectories and resolution by pitting out-group aggression against in-group defense. Because existing models of intergroup conflict such as *N*-person prisoners' dilemmas and intergroup contest games are ill-fitted to distinguish between out-group aggression and in-group defense, we developed an intergroup aggressor–defender conflict (IADC) game. Six individuals randomly divided into three-person aggressor and defender groups each received 20 Experimental Euros from which they could contribute g ($0 \leq g_i \leq 20$) to their group's pool C ($0 \leq C \leq 60$). Individual contributions to the pool were wasted, but when $C_{aggressor} > C_{defender}$, the aggressor won the remaining resources of the defenders ($60 - C_{defender}$), which was divided equally among aggressor group members and added to their remaining endowments ($20 - g_i$). Defenders thus earned 0 when aggressors won. However, when $C_{aggressor} \leq C_{defender}$, defenders survived, and individuals on both sides kept their remaining endowments ($20 - g_i$). Thus, individual contributions in aggressor (defender) groups reflect out-group aggression (in-group defense). We used the game to test whether individual contributions to out-group aggression are weaker than those to in-group defense, examine how this possible difference translates into aggressor's success in subordinating its defender, and determine whether possible failures to subordinate defender groups are the result of a lack of motivation to contribute to out-group aggression and/or to a failure to align and coordinate individual contributions to out-group aggression.

Method Summary

The IADC was implemented in two experiments. In Exp. 1, $n = 144$ subjects participated (106 females; median age, 21 y). In Exp. 2, $n = 132$ subjects participated (78 females; median age, 22 y). In each experiment, one session involved six subjects divided at random into a three-person aggressor and a three-person defender group; Exp. 1 thus has 24 (144/6) IADC sessions, and Exp. 2 had 22 (132/6) IADC sessions. In both experiments, the six individuals invited for a single IADC session were randomly assigned to one of two laboratory rooms and one of three individual cubicles within that room. Subjects were unaware of who else was in either laboratory room and, once seated, signed informed consent and read instructions for the IADC (*Materials and Methods*). Thereafter, subjects indicated their contribution g ($0 \leq g_i \leq 20$) to their group's pool C and were informed about the total contribution their group made to C ($0 \leq C \leq 60$), the total contribution C made by the other group, and the resulting earnings to the members of their own group, themselves included. This feedback concluded one IADC episode. In total, subjects engaged in one block of five baseline episodes and one block of five treatment episodes (i.e., allowing for peer punishment in Exp. 1 and for sequential decision-making in Exp. 2; further detail follows). The order in which blocks were presented was counter balanced and found not to qualify the conclusions drawn here.

Investments were always wasted, and, from a social welfare perspective, it thus is optimal for all individuals on both sides not to contribute anything. This social welfare perspective contrasts with both individual and group welfare considerations. Specifically, the IADC has mixed-strategy Nash equilibria in which individuals contribute to out-group aggression (versus in-group defense) on average mean = 10.15 (versus mean = 9.77). This analysis also implies that aggressor (versus defender) groups win (versus survive) 32.45% (versus 67.55%) of the episodes (35) (*Materials and Methods*). We examined these estimates against the data from the five baseline episodes of the two experiments combined ($n = 276$ individuals in 46 IADCs). Out-group aggression fell below (mean = −2.401; SE = 0.567), and in-group defense exceeded (mean = 0.858; SE = 0.400), the Nash equilibrium [$t(45) = -9.231$ ($P \leq 0.001$) and $t(45) = 2.146$ ($P = 0.037$)]. Aggressors defeated defenders in 22.5% of their attacks, which is below the Nash success rate [mean = −0.679; SE = 0.154; $t(45) = -4.405$; $P \leq 0.001$].

Experiment 1. As noted, a first possible explanation for the relatively low success rate for out-group aggression is a relatively low willingness to contribute to the aggressor's fighting capacity. If true, sanctioning arrangements that are known to increase contributions to public goods should increase contributions more in aggressor groups than in defender groups (in which contributions are already high). If sanctions indeed affect contributions, especially in aggressor groups, and if relatively low willingness to invest is a cause for the aggressor's low success rate, sanctions may also increase the aggressor group's success rate.

One sanctioning arrangement that can increase costly contributions is peer punishment. Individuals, after they see their group members' contributions, can execute a punishment that is costly to themselves, but more costly to the punished group member or members (13, 19, 36–39). Experiments have shown that individuals punish to motivate others to contribute more and that individuals respond to (the threat of) punishment by increasing subsequent contributions in public good provision (36–39) and intergroup contests (13, 18, 19). Accordingly, Exp. 1 examined whether, relative to baseline episodes in which peer punishment was absent, the presence of peer punishment increased contributions to the group's fighting capacity, especially in aggressor groups, and whether such relative increase in out-group aggression translates into higher success rates for aggressor groups. The experiment involved five baseline episodes and five consecutive episodes in which individuals could assign costly punishment within groups. In episodes with peer punishment, each player i received 10 "decrement points" and could assign s ($0 \leq s_{i,j} \leq 5$) to any other player j in their group, with each point assigned reducing 1 point from the punisher i's Experimental Euros (EE), and 3 points from the punished player j's EE (punishment across groups was not possible). As in baseline episodes, resulting earnings were then shown, which ended the episode [on each round, we randomly reshuffled the letter by which group members were identified, so that within the group, (expecting) punishment was decoupled from reputation and reciprocity considerations].

Data were aggregated to the group level and submitted to a 2 (role: aggressor/defender) × 2 (punishment: present/absent) ANOVA. Contributions to in-group defense were higher than to out-group aggression [$F(1, 23) = 41.97$; $P = 0.0001$]. Importantly, punishment increased contributions to out-group aggression [$F(1, 23) = 4.49$; $P = 0.046$], but not to in-group defense [$F(1, 23) = 1.18$; $P = 0.289$] (Fig. 1*A*). Reflecting less coordination in aggressor groups, we observed that within-group dispersion in a conflict episode was larger for out-group aggression than for in-group defense [$F(1, 23) = 14.52$; $P = 0.001$], and dispersion was not influenced by punishment [Fig. 1*B*; role × punishment: $F(1, 23) = 1.26$; $P = 0.276$]. Zooming in on noncontributors (individuals who invested zero, within groups and across episodes), ANOVA revealed effects for role [$F(1, 23) = 21.22$; $P = 0.001$], punishment [$F(1, 23) = 9.25$; $P = 0.006$], and role × punishment [$F(1, 23) = 8.60$; $P = 0.008$] (Fig. 1*C*). Punishment did not affect the (very low) number of people not contributing to in-group defense, but reduced the higher number of people not contributing to out-group aggression from 23% to 13%. Thus, peer punishment

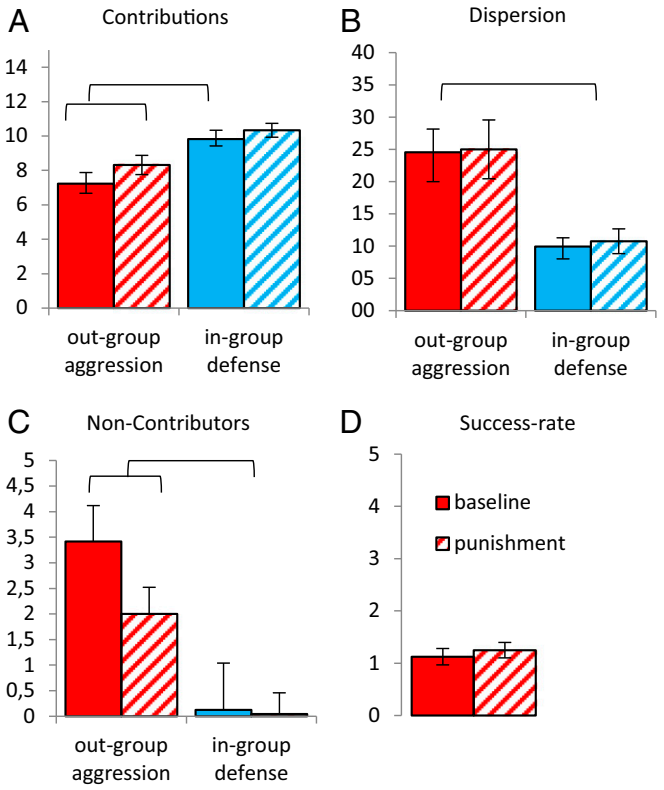

Fig. 1. Peer punishment in intergroup aggressor–defender conflict (displayed mean ± 1 SE). Connectors indicate difference at $P \leq 0.05$. (*A*) Contributions (range, 0–20). (*B*) Within-group variance (dispersion). (*C*) Number of non-contributors per group across conflict episodes (range, 0–5). (*D*) Aggressor success (range, 0–5).

increased out-group aggression more than in-group defense. This increased motivation notwithstanding, punishment failed to increase success: Aggressor groups only won 23.75% of all episodes, a success rate not conditioned by punishment [$F(1, 23) \leq 0.35$; all $P \geq 0.588$] (Fig. 1*D*).

In Exp. 1, peer punishment increased contributions more in aggressor than defender groups, but the increased fighting capacity in aggressor groups did not increase success (and reduced individual wealth; *Materials and Methods*). The relatively low success rate for out-group aggression cannot be simply elevated by increasing the contributions. Exp. 2 targeted the alternative possibility: out-group aggression fails because of poor coordination. If true, arrangements that enable groups to align their contributions into coordinated fighting should be particularly effective in aggressor groups, thus increasing their success rate. One such arrangement is sequential decision-making (40, 41, 51), which has been shown to solve collective action problems in public goods provision (40–43). In such a procedure, one individual moves first, allowing the rest of the group to adapt and follow the first-mover's lead (40, 41, 43). It is seen in group-hunting carnivores such as wolves (upon encircling their prey, the group waits until the most senior wolf leads by launching the first attack) (25, 44), and has been identified as a minimal form of leadership with voluntary followers (45, 46).

Experiment 2. In addition to the five baseline (simultaneous decision-making) episodes, Exp. 2 included five episodes of sequential decision-making: one member in each group was randomly selected to move first, then the randomly selected second player made their decision, and then the remaining third player made their decision (43). Each decision was shown to the other two group members. The episode ended with back-reporting earnings.

Data were submitted to a 2 (role: aggressor/defender) × 2 (decision-making procedure: simultaneous/sequential) mixed-model ANOVA. Contributions to in-group defense were higher than to out-group aggression [$F(1, 21) = 29.30$; $P \leq 0.001$) and were not affected by decision-making procedure [$F(1, 21) = 0.07$; $P = 0.799$] or the role × procedure interaction [$F(1, 21) = 2.71$; $P = 0.115$] (Fig. 2*A*). As in Exp. 1, dispersion was larger for out-group aggression than for in-group defense [$F(1, 21) = 5.42$; $P = 0.030$]. However, a role × procedure interaction [$F(1, 21) = 5.04$; $P = 0.036$) showed that sequential decision-making reduced within-episode dispersion for out-group aggression, but not for in-group defense (Fig. 2*B*). Zooming in on noncontributors, ANOVA revealed effects for role [$F(1, 21) = 17.52$; $P \leq 0.001$] and role × procedure [$F(1, 21) = 6.36$; $P = 0.020$] (Fig. 2*C*). Sequential decision-making did not affect the (low) number of people not contributing to in-group defense; in aggressor groups, however, sequential decision-making reduced the (higher) number of people not contributing to out-group aggression from 31% to 23%. Crucially, sequential decision-making almost doubled the aggressor's success, from 20% under simultaneous decision-making to 35% under sequential decision-making [$F(1, 21) = 6.05$; $P = 0.023$) (Fig. 2*D*).

Conclusions and Discussion

The experiments together showed that individual contributions to out-group aggression are weaker than those to in-group defense, and aggressor groups frequently fail to win the conflict and waste individual resources on ineffective out-group aggression. This failure is unlikely to be caused by a lack of motivation to contribute to out-group aggression. Exp. 1 showed that peer punishment motivated individuals to contribute more to out-group aggression

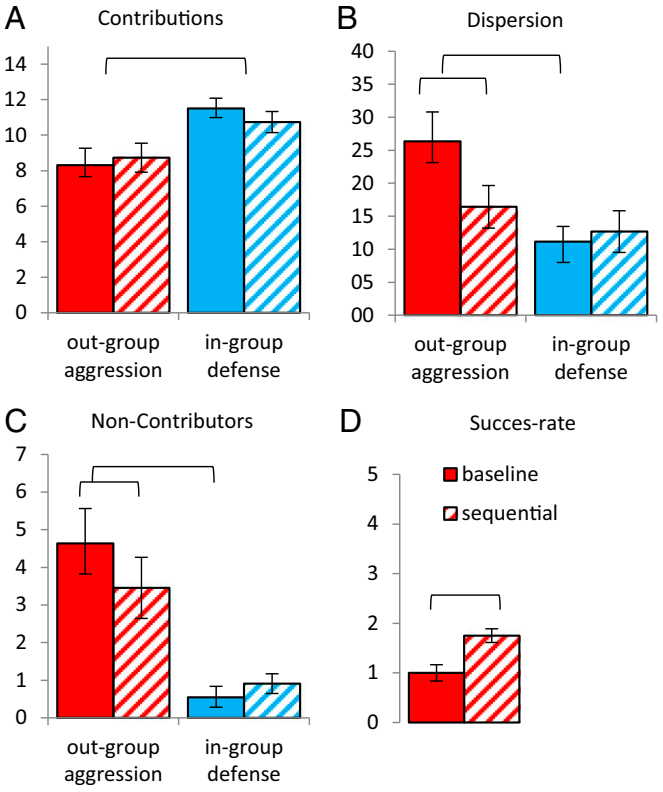

Fig. 2. Sequential decision-making in intergroup aggressor–defender conflict (displayed mean ± 1 SE). Connectors indicate difference at $P \leq 0.05$. (*A*) Contributions (range, 0–20). (*B*) Within-group variance (dispersion). (*C*) Number of noncontributors per group across conflict episodes (range, 0–5). (*D*) Aggressor success (range, 0–5).

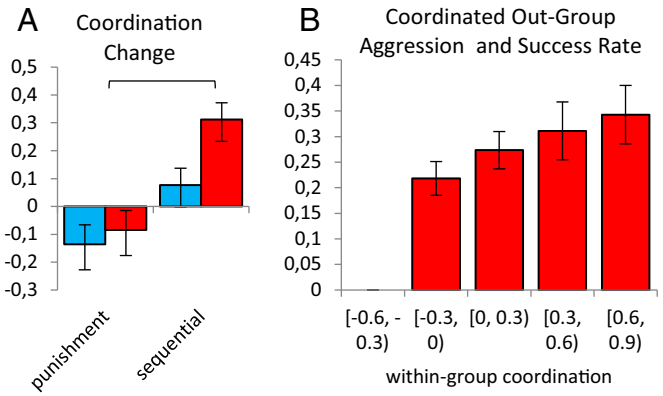

Fig. 3. Coordination in intergroup aggressor–defender conflict (displayed mean intraclass correlation ± 1 SE). Connectors indicate difference at $P \leq 0.05$. (*A*) Change from baseline when punishment or sequential decision-making is introduced. Blue (red) bars are defender (aggressor) groups. (*B*) Aggressor success as a function of aggressor's within-group coordination.

(but not to in-group defense), yet such higher contributions did not translate into increased success rates for out-group aggression, leading to more wasted resources and lower overall welfare.

Exp. 2 suggested that the relatively low success rate for aggressor groups can be attributed to a failure to align and coordinate individual contributions to out-group aggression into effective collective action. This possibility was tested directly by computing, as an index of coordination, the within-episode intraclass correlation for contributions (47) (*Materials and Methods*). Relative to baseline, sequential decision-making increased coordination in aggressor groups more than in defender groups (Fig. 3*A*). Also, as shown, sequential decision-making improved coordination more than peer punishment, and coordination predicted success for out-group aggression [$r = 0.30$; $t(90) = 2.94$; $P = 0.004$; Fig. 3*B*]. It follows that the aggressor group's failure to subordinate its defender is a result of the aggressor's tougher task of coordinating within-group contributions into effective out-group aggression.

Willingness to contribute, coordinated collective action, and aggressor success rates were revealed in an intergroup conflict that modeled a clashing of out-group aggression by one antagonist and in-group defense by its opponent. Real-world analogies are group-hunting carnivores facing prey aggressively defending themselves, boards of directors attempting and warding off a hostile takeover, tribal raiding and warfare, and most interstate disputes. For example, of the 2,209 documented interstate conflicts since the Congress of Vienna in 1816 (27, 48), 67% were between aggressors seeking territorial or policy change in states that tried to defend the status quo (*Materials and Methods*). Similar to our model, these aggressor–defender conflicts typically see an aggressor success rate of around 35%: aggressor states win less than 30% of the interstate conflicts in which they are involved, and industry boards pushing for hostile takeover are successful only 40% of the time (Fig. 4*A*) (49–51) (*Materials and Methods*). Even hunting groups of wolves, lions, jackals, or killer whales are successful once in every three attempts (33%; Fig. 4*B*) (24, 44, 52–58) (*Materials and Methods*).

The finding that, across species and types of intergroup conflict, aggressors succeed a third of the time on average may be a result of the need to coordinate collective action into a costly attack sometimes, but not all of the time. Indeed, aggressing all of the time is energetically impossible. Also, it would set a permanent high level of in-group defense and prohibit defender groups from being lured into an illusionary state of safety, with lowered defense and concomitant higher probability of successful capture (31). To trump in-group defense, aggressors need to launch surprise attacks. Next to a willingness to sacrifice private resources, launching surprise attacks requires careful within-group coordination.

Our conclusions derive, in part, from two laboratory experiments and may be limited to the specific parameters used to design the IADC. In many intergroup conflicts, including those analyzed here, a single failure to defend adequately will result in the death for the prey, yet after a failure to capture, a predator can find an alternative prey. As noted, however, attacking is very costly, and when a predator repeatedly fails on consecutive attacks, it dies just like the prey that fails to adequately defend. Similarly, a company attempting but failing a hostile takeover may be weakened to the extent that bankruptcy cannot be avoided. Thus, whereas in the current experiments both aggressor and defender groups received a full reset of their endowments on each new round, oftentimes such a reset can be less abundant or substantially delayed, and the cost of unsuccessful attack may be (much) higher than in our experiments. Whether these deter individuals from contributing to out-group aggression or stimulate contributions and facilitate coordination of collective action remains an issue for further research.

It has been argued that histories of intergroup conflict and competition may have acted as selection pressures favoring self-sacrificial contributions to one's group's fighting capacity and contributed to the development and spread of institutions and technologies

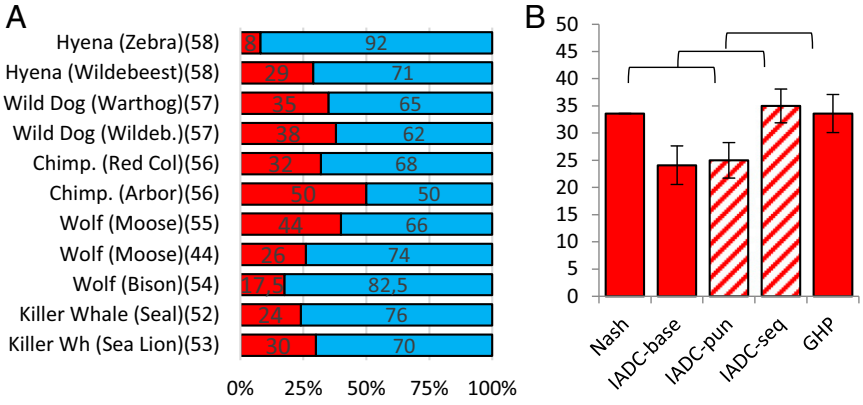

Fig. 4. Aggressor-defender success-rates. (*A*) Percentage of successful attacks by group-hunting animals (and their prey). Red (blue) bars are predator (prey). Numbers in bars are observed cases; bracketed numbers in *y* axis are source references. (*B*) Nash estimate for aggressor's success in the IADC (Nash), observed aggressor success in baseline treatments of Exp. 1 and Exp. 2 (IADC-base), punishment (IADC-pun), and sequential decision-making (IADC-seq), and sample-size weighted average success rate in group hunting predators (GHP) (displayed percentage ± 1 SE); connectors indicate difference at $P \leq 0.05$.

that enable groups to coordinate their members' activities and contributions (3, 14). Current findings align with these possibilities. However, the relatively high success rate of in-group defense suggests that evolutionary and cultural pressures may have favored capacities for cooperation and coordination when the group goal is to defend, rather than to expand, dominate, and exploit.

Materials and Methods

Experiments were approved by the University of Amsterdam Psychology Research Ethics Board (files 2014-WOP-3451 and 2015-WOP-4531); subjects provided written informed consent before the experiment and were debriefed. Subjects were recruited on the university campus through an online recruiting website for a study announced as "human decision making in groups." The experimental instructions used neutral language throughout (e.g., groups were referred to as group A and group B, contributions were labeled investments, and terms such as in-group defense and out-group aggression were avoided). All subjects passed a comprehension check that consisted of two complete scenarios for one episode of the IADC from the perspective of their role, with their group winning and losing the episode, respectively. Experiments involved no deception, and subjects received a €10 show-up fee and mean = €3.62 (range, 0–€10) for their performance. Personal earnings in both experiments were based on the average of two randomly selected baseline episodes and two punishment (Exp. 1) or sequential decision-making (Exp. 2) episodes, provided that earnings would not drop below the €10 show-up fee and that both groups were rewarded equally (per local policies within our research laboratories). To preserve confidentiality, earnings were calculated afterward and transferred to the subject's bank account.

Game-Theoretic Analysis. Game-theoretic equilibria for the IADC game, with two three-person groups, each member assumed to have risk-neutral preferences, and a discretionary resource to invest from, were numerically estimated using a modified version of an algorithm developed by Chatterjee (35) in Matlab. The resulting unique mixed-strategy Nash equilibrium assigns the same strategy for players within the same group. For each pure strategy (range, 0–20), the probabilities for investing in out-group aggression (in-group defense) are $P(0) = 0.5322$ (0.0105), $P(1) = 0.0876$ (0.5615), $P(2) = 0.045$ (0.1050), $P(3) = 0.0321$ (0.0249), $P(4) = 0.0068$ (0.0241), $P(5) = 0.0067$ (0.0198), $P(6) = 0.0095$ (0.0894), $P(7) = 0.0283$ (0.0844), $P(8) = 0.1125$ (0.0087), $P(9) = 0.0152$ (0.0076), $P(10) = 0.0066$ (0.0067), $P(11) = 0.0054$ (0.0051), $P(12) = 0.0046$ (0.0044), $P(13) = 0.0054$ (0.0050), $P(14) = 0.0134$ (0.0064), $P(15) = 0.0594$ (0.0080), $P(16) = 0.0147$ (0.0089), $P(17) = 0.0043$ (0.0073), $P(18) = 0.0024$ (0.0053), $P(19) = 0.0019$ (0.0040), and $P(20) = 0.0015$ (0.0031). Thus, assuming common belief in rationality in individual group members, out-group aggression (in-group defense) is expected to average 10.15 (9.77), and aggressors (defenders) should win (survive) 32.45% (67.55%) of the episodes.

An alternative approach is to treat groups as single agents, with each group having risk-neutral preferences and being endowed with $20 \times 3 = 60$ resources. The strategies played in equilibrium imply that both groups only assign positive probabilities to strategies between 0 and 38 (i.e,, ref. 30). This approach yields expected out-group aggression (in-group defense) of 5.41 (7.25), and aggressors (defenders) should win (survive) 37.51% (62.49%) of the episodes. These estimates differ more from observed contributions and success rates than those predicted by the admittedly more realistic individual-level equilibria.

Indexing Within-Group Coordination. The intraclass correlation [(ICC(2)] describes how strongly individuals in the same group resemble each other. Unlike most other correlation measures, it operates on data structured as groups, rather than data structured as paired observations. The index can be used to assess the amount of statistical interdependence within a particular social system (e.g., work-team) underlying individual-level data (e.g., individual ratings of group cohesion). Higher ICC(2) values reflect the level of consensus + consistency one would expect if an individual contributor was randomly selected from his or her group and within a particular decision round, and his or her scores were compared with the mean score (i.e., estimated true score) obtained from this group (47). Thus, higher ICC(2) values in essence mean group members are more similar to each other in the contributions made to their group's fighting capacity.

Additional Results. In both experiments, we explored the influence of conflict episode in 2 (role) × 2 (treatment) × 5 (episode) ANOVAs. In Exp. 1, we found no effects involving episode, all Fs < 1.28, all Ps > 0.25. In Exp. 2, we found that the role × sequence effect on dispersion (Fig. 2*B*) was qualified by a role × sequence × episode effect [$F(4,18) = 4.736$; $P = 0.009$]. The lower dispersion in aggressor groups under sequential decision-making disappeared in the final episode, which may reflect an end-game effect. We suggest that our main conclusions hold across conflict episodes.

In Exp. 1, we looked at targets of punishment. We identified weak contributors ($g \leq 5$) receiving punishment ("weak contributors punished") or not ("weak contributors not punished"), and strong contributors ($g \geq 15$) receiving punishment ("strong contributors punished") or not ("strong contributors not punished"). A 2 (role) × 2 (contributor type: weak/strong) × 2 (contributor type punished: yes/no) within-session ANOVA showed that in aggressor groups, more weak than strong contributors were punished [mean = 3.0 vs. mean = 1.2; $F(1, 23) = 10.33$; $P = 0.005$), whereas in defender groups, both types were equally unlikely to receive punishment [mean = 1.10 vs. mean = 1.24; $F(1, 23) = 0.02$; $P = 0.890$]. Thus, in particular, aggressor groups biased punishment toward their weak contributors.

In both experiments, we examined individual wealth as a function of treatment and role. Intergroup conflict is wasteful, which the experimental game mirrored. Investments were always wasted, and individuals in defender (aggressor) groups could earn between 0 and 20 EE (0 and 40 EE). Despite these differences in stakes, however, individuals in aggressor (defender) groups lost about 30% (35%) of their individual wealth (final wealth/20 EE). In Exp. 1, we observed effects for role [$F(1, 22) = 289.53$; $P \leq 0.0001$] and punishment [$F(1, 22) = 3.32$; $P = 0.081$] (marginal). Individuals in aggressor groups experienced a greater loss in wealth under punishment (mean = 14.206 vs. mean = 15.317), as did individuals in defender groups (mean = 7.111 vs. mean = 7.633). These numbers are conservative estimates because they ignore wealth reductions resulting from punishing others and being punished. In Exp. 2, we found that wealth was affected by both role [$F(1, 21) = 254.13$; $P \leq 0.001$] and role × decision-making procedure [$F(1, 21) = 7.91$; $P = 0.010$]. Under sequential decision-making, individuals in aggressor groups saw less wealth reduction than in baseline conditions [mean = 14.803 (SE = 0.609) vs mean = 13.469 (SE = 0.806)]; individuals in defender groups lost more under sequential decision-making [mean = 6.712 (SE = 0.654) vs. mean = 5.724 (SE = 0.649)], which is a direct consequence of their aggressors becoming more effective under sequential decision-making (Fig. 2*D*). Thus, in aggressor groups, the introduction of peer punishment reduced, and sequential decision-making increased, wealth.

Because individuals were randomly assigned to groups, we had all-male, all-female, and mixed-sex groups. A meta-analysis (16) found no significant differences between male and female participants in costly contributions to in-group efficiency or out-group competitiveness. The absence of significant sex differences was replicated here: Across current experiments, correlations among group-level contributions, within-group dispersion, and success-rate for in-group defense and out-group aggression on the one hand, and the number of males in aggressor and defender groups on the other, ranged between −0.251 and +0.112, with all Ps ≥ 0.10. Current findings and conclusions generalize across sex and group composition, and we suggest that contributing to the group's fighting capacity may not be sex-specific.

Archival Analyses: Interstate Conflict, Hostile Takeovers, and Group-Hunting Predators. The Correlates of War project provides descriptive information on 2,586 interstate (militarized) conflicts since the Congress of Vienna in 1816 (27, 48). We integrated distinct datasets (MIDA and MIDB; versions 4.01; both downloaded July 15, 2014, from www.correlatesofwar.org) to determine the structure of the interstate conflict as being symmetrical (0 = between two aggressor states or between two defender states) or asymmetrical (1 = between an aggressor and a defender state). States are "revisionist" (aggressor) when they desire change in territory, policy, or government in their antagonist; nonrevisionists (defenders), in contrast, seek to preserve and maintain the status quo with regard to territory, policy, or government (27, 48). Exactly two-thirds (67%) were between an aggressor and a defender state, and 33% were symmetrical ($\chi^2[1,2209] = 494.45$; $P \leq 0.0001$). The datasets also contained coding for the outcome of these aggressor–defender disputes: aggressors were unsuccessful in 1,057 disputes (985 ended in a stalemate and 72 ended in victory to the defender). Aggressors were relatively victorious in 239 disputes, reaching either a compromise (76) or a clear victory (163). Two-hundred sixty cases were coded "unclear." Excluding these gives a conservative estimate of aggressor success of 18%; coding "unclear" as aggressor success gives a liberal 38%, with the point estimate thus being 28% (see also Fig. 4*B*).

After a survey of the literature on hostile takeover (26), we retained three sources that provided sufficient statistical detail on the number of hostile takeovers that were or were not successful. Takeover attempts were defined as hostile when the target firm (defender) officially rejected an offer but the acquirer (aggressor) persisted with the takeover (26), and thus represent a

clashing of out-group aggression and in-group defense (e.g., the use of "poison pills"). Success was coded as takeover completed (1) or abandoned (0). Mitchell and Mulherin (50) analyzed takeover activity by major industrial corporations between 1982 and 1989. Takeover attempts considered friendly were successful in 268 of 286 documented cases (93.7%); Takeover attempts considered hostile were successful in 85 out of the 243 documented cases (35.0%). Scheper and Guillen (49) collected data on 37 countries between 1988 and 1998 and detected 952 hostile takeover attempts, of which 336 were coded as successful (35.3%). Secondary analyses on data from Muehfeld, Sabib, and Van Witteloostuijn (51), who examined takeover activity in the newspaper industry between 1981 and 2000, revealed that 3,173 of the 3,615 cases were coded friendly and 442 as hostile. Completion rate was 76% for friendly and 53.2% for hostile takeovers (235/442). This figure is higher than those reported in refs. 49 and 50, possibly because these other sources considered mostly publicly listed companies with often sophisticated measures against hostile takeovers (e.g., "poison pills"). Such measures may be less developed or even absent altogether in the smaller companies present in the data from (51), and the lack of defense mechanisms may explain the higher success rate seen for hostile takeovers. Notwithstanding the variability in years of study, type of industry, and geopolitical regions, the sample size weighted success rate for hostile takeovers averages 40.1% (656/1,637).

Success rates for group-hunting predators were obtained by tracking citations to refs. 24 and 25; surveying Web of Science (Nov. 2015), using the search terms "group" (or "collective") AND "hunting" (or "predation;" "predators;" "carnivores") AND "success" (or "kills;" "attacks;" "killings", "prey capture;") and tracking citations to articles obtained under the first two methods. Included in the analysis here are reports focusing on mammalian predators with prey fighting back as the dominant response (rather than fleeing) and providing sufficient statistical detail to obtain a reliable estimate of predator success. Retained are refs. 44 and 52–58.

ACKNOWLEDGMENTS. We thank Mathijs van Dijk, Brian Burgoon, and Katrin Muehlfeld for their help at various stages of this project. Financial support was provided by Netherlands Science Foundation Grant 432.08.002 and the Netherlands Institute for Advanced Study.

1. Macfarlan SJ, Walker RS, Flinn MV, Chagnon NA (2014) Lethal coalitionary aggression and long-term alliance formation among Yanomamö men. *Proc Natl Acad Sci USA* 111(47):16662–16669.
2. Whitehouse H, McQuinn B, Buhrmester M, Swann WB, Jr (2014) Brothers in arms: Libyan revolutionaries bond like family. *Proc Natl Acad Sci USA* 111(50):17783–17785.
3. Choi J-K, Bowles S (2007) The coevolution of parochial altruism and war. *Science* 318(5850):636–640.
4. Bornstein G (2003) Intergroup conflict: Individual, group, and collective interests. *Pers Soc Psychol Rev* 7(2):129–145.
5. Yamagishi T, Mifune N (2016) Parochial altruism: Does it explain modern human group psychology? *Curr Opin Psychiatry* 7:39–43.
6. Bar-Tal D, Halperin E, de Rivera J (2007) Collective emotions in conflict situations: Societal implications. *J Soc Issues* 63(2):441–460.
7. Holbrooke R (1998) *To End a War* (Random House, New York).
8. Staub E (1996) Cultural-societal roots of violence. The examples of genocidal violence and of contemporary youth violence in the United States. *Am Psychol* 51(2):117–132.
9. Sternberg RJ (2003) A duplex theory of hate: Development and application to terrorism, massacres, and genocide. *Rev Gen Psychol* 7(3):299–328.
10. Christ O, et al. (2014) Contextual effect of positive intergroup contact on outgroup prejudice. *Proc Natl Acad Sci USA* 111(11):3996–4000.
11. Manson JH, Wrangham RW (1991) Intergroup aggression in chimpanzees. *Curr Antrhop* 32(4):369–390.
12. Wilson DS, Wilson EO (2007) Rethinking the theoretical foundation of sociobiology. *Q Rev Biol* 82(4):327–348.
13. Mathew S, Boyd R (2011) Punishment sustains large-scale cooperation in prestate warfare. *Proc Natl Acad Sci USA* 108(28):11375–11380.
14. Boyd R, Richerson PJ (1982) Cultural transmission and the evolution of cooperation. *Hum Ecol* 10:325.
15. Wildschut T, Pinter B, Vevea JL, Insko CA, Schopler J (2003) Beyond the group mind: A quantitative review of the interindividual-intergroup discontinuity effect. *Psychol Bull* 129(5):698–722.
16. Balliet D, Wu J, De Dreu CKW (2014) Ingroup favoritism in cooperation: A meta-analysis. *Psychol Bull* 140(6):1556–1581.
17. Rapoport A, Bornstein G (1987) Intergroup competition for the provision of binary public goods. *Psychol Rev* 94(3):291–299.
18. Abbink K, et al. (2012) Parochial altruism in inter-group conflicts. *Econ Lett* 117(1): 45–48.
19. Abbink K, Brandts J, Herrmann B, Orzen H (2010) Inter-group conflict and intra-group punishment in an experimental contest game. *Am Econ Rev* 100(1):420–447.
20. Lacomba JA, Lagos F, Reuben E, Van Winden F (2014) On the escalation and de-escalation of conflict. *Games Econ Behav* 86:40–57.
21. Dechenaux E, Kovenock D, Sheremeta RM (2015) A survey of experimental research on contests, all-pay auctions, and tournaments. *Exp Econ* 18(4):609–669.
22. Waytz A, Young LL, Ginges J (2014) Motive attribution asymmetry for love vs. hate drives intractable conflict. *Proc Natl Acad Sci USA* 111(44):15687–15692.
23. De Dreu CKW, et al. (2010) The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* 328(5984):1408–1411.
24. Packer C, Ruttan L (1988) The evolution of cooperative hunting. *Am Nat* 132:159–198.
25. Bailey I, et al. (2013) Group hunting within the Carnivora: Physiological, cognitive and environmental influences on strategy and cooperation. *Behav Ecol Sociobiol* 67:1–17.
26. Schwert GW (2000) Hostility in takeovers: In the eyes of the beholder? *J Finance* 55(6): 2599–2640.
27. Jones DM, et al. (1996) Militarized interstate disputes 1816-1992: Rationale, coding rules, and empirical patterns. *Confl Mngmt Peace Science* 15(2):163–213.
28. Hamilton WD (1971) Geometry for the selfish herd. *J Theor Biol* 31(2):295–311.
29. Nelson RJ, Trainor BC (2007) Neural mechanisms of aggression. *Nat Rev Neurosci* 8(7): 536–546.
30. De Dreu CKW, Scholte HS, van Winden FA, Ridderinkhof KR (2015) Oxytocin tempers calculated greed but not impulsive defense in predator-prey contests. *Soc Cogn Affect Neurosci* 10(5):721–728.
31. De Dreu CKW, Kret ME, Sligte IG (April 1, 2016) Modulating prefrontal control in humans reveals distinct pathways to competitive success and collective waste. *Soc Cogn Affect Neurosci*, 10.1093/scan/nsw045.
32. Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.
33. Simunovic D, et al. (2013) Preemptive strike: An experimental study of fear-based aggression. *J Exp Soc Psychol* 49(6):1120–1123.
34. Rusch H (2013) Asymmetries in altruistic behavior during violent intergroup conflict. *Evol Psychol* 11(5):973–993.
35. Chatterjee B (2009) An optimization formulation to compute Nash equilibrium in finite games. *Proceedings of International Conference on Methods and Models in Computer Science* (IEEE, Piscataway, NJ), pp. 1–5.
36. Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415(6868):137–140.
37. Egas M, Riedl A (2008) The economics of altruistic punishment and the maintenance of cooperation. *Proc Biol Sci* 275(1637):871–878.
38. Balliet D, Mulder LB, Van Lange PAM (2011) Reward, punishment, and cooperation: A meta-analysis. *Psychol Bull* 137(4):594–615.
39. Gross J, Méder ZZ, Okamoto-Barth S, Riedl A (2016) Building the Leviathan–Voluntary centralisation of punishment power sustains cooperation in humans. *Sci Rep* 6:20767.
40. Levati MV, et al. (2007) Leading by example in a public goods experiment with heterogeneity and incomplete information. *J Conflict Resolut* 51(5):793–818.
41. Glowacki L, von Rueden C (2015) Leadership solves collective action problems in small-scale societies. *Philos Trans R Soc Lond B Biol Sci* 370(1683):20150010.
42. Gavrilets S, Fortunato L (2014) A solution to the collective action problem in between-group conflict with within-group inequality. *Nat Commun* 5:3526.
43. Budescu DV, Au WT, Chen XP (1997) Effects of protocol of play and social orientation on behavior in sequential resource dilemmas. *Organ Behav Hum Decis Process* 69(3): 179–193.
44. Wikenros C, Sand H, Wabakken P, Liberg O, Pedersen HC (2009) Wolf predation on moose and roe deer: Chase distances and outcome of encounters. *Acta Theriol (Warsz)* 54(3):207–218.
45. Hermalin BE (1998) Toward an economic theory of leadership: Leading by example. *Am Econ Rev* 88(5):1188–1206.
46. Potters J, Sefton M, Vesterlund L (2007) Leading-by-example and signaling in voluntary contribution games: An experimental study. *Econ Theory* 33(1):169–182.
47. LeBreton JM, Senter JL (2008) Answers to 20 questions about interrater reliability and interrater agreement. *Organ Res Methods* 11(4):815–852.
48. Wright TM (2014) Territorial revision and state repression. *J Peace Res* 51(3):375–387.
49. Scheper WD, Guillen MF (2004) Stakeholder rights and corporate governance: A cross-national study of hostile takeovers. *Adm Sci Q* 49(2):263–295.
50. Mitchell ML, Mulherin JH (1996) The impact of industry shocks on takeover and restructuring activity. *J Financ Econ* 41(2):193–229.
51. Muehfeld K, Sabib PR, Van Witteloostuijn A (2007) Completion or abandonment of mergers and acquisitions: Evidence from the newspaper industry, 1981–2000. *J Media Econ* 20(2):107–137.
52. Hoelzel AR (1991) Killer whale predation on marine mammals at Punta Norte, Argentina. *Behav Ecol Sociobiol* 29(3):197–204.
53. Pitman RL, Durban JW (2012) Cooperative hunting behavior, prey selectivity and prey handling by lack ice killer whales (*Orcinus orca*), type B, in Antarctic Peninsula waters. *Mar Mamm Sci* 28:16–36.
54. MacNulty DR, Tallian A, Stahler DR, Smith DW (2014) Influence of group size on the success of wolves hunting bison. *PLoS One* 9(11):e112884.
55. Sand H, Wikenros C, Wabakken P, Liberg O (2006) Effects of hunting group size, snow depth and age on the success of wolves hunting moose. *Anim Behav* 72(4):781–789.
56. Boesch Ch (1994) Cooperative hunting in wild chimpanzees. *Anim Behav* 48(3): 653–667.
57. Creel S, Creel NM (1995) Communal hunting and pack size in African wild dogs, *Lycaon pictus*. *Anim Behav* 50(5):1325–1339.
58. Holekamp KE, Smale L, Berg R, Cooper SM (1997) Hunting rates and hunting success in the spotted hyena (*Crocuta crocuta*). *J Zool (Lond)* 242(1):1–15.

Understanding adolescence as a period of social–affective engagement and goal flexibility

*Eveline A. Crone[1] * and Ronald E. Dahl[2] **

Abstract | Research has demonstrated that extensive structural and functional brain development continues throughout adolescence. A popular notion emerging from this work states that a relative immaturity in frontal cortical neural systems could explain adolescents' high rates of risk-taking, substance use and other dangerous behaviours. However, developmental neuroimaging studies do not support a simple model of frontal cortical immaturity. Rather, growing evidence points to the importance of changes in social and affective processing, which begin around the onset of puberty, as crucial to understanding these adolescent vulnerabilities. These changes in social–affective processing also may confer some adaptive advantages, such as greater flexibility in adjusting one's intrinsic motivations and goal priorities amidst changing social contexts in adolescence.

Cognitive control
A set of neurocognitive processes that are important for achieving short- and long-term goals, particularly when individuals are required to adjust their thoughts and actions adaptively in response to changing environmental demands in order to achieve their goal.

[1]*Department of Psychology, Leiden University, Wassenaarseweg 52, 2333AK, The Netherlands.*
[2]*Institute of Human Development, University of California, Berkeley, California 94707, USA.*
e-mails: rondahl@berkeley.edu; ECrone@FSW.leidenuniv.nl
**All authors contributed equally to this work.*
doi:10.1038/nrn3313

Adolescence, which is defined as the transition phase between childhood and adulthood, is a natural time of learning and adjustment, particularly in the setting of long-term goals and personal aspirations (BOX 1). It also is a time when youths are discovering how to navigate new, often compelling, social challenges and are adjusting to myriad physical, cognitive and emotional changes within themselves[1,2]. The onset of adolescence is characterized by the start of pubertal maturation, which typically begins between 9 and 12 years of age (usually 1–2 years earlier in girls than in boys). The onset of puberty creates a cascade of hormonal changes — including dramatic increases in the secretion of adrenal androgens, gonadal steroids and growth hormone (BOX 2). This surge in hormones has a central role within a larger set of biological changes in the process of achieving reproductive maturity. These changes include: rapid physical growth; sexually dimorphic alterations in facial structure, voice and body characteristics; metabolic changes; the activation of new drives and motivations; changes in sleep and circadian regulation; and a wide array of social, behavioural and emotional changes[3].

Although the beginning of adolescence is characterized by distinct and dramatic physiological changes, the end of adolescence has less clear biological boundaries. Attaining 'adulthood' involves changes in social roles and responsibilities, is partly culturally defined and typically extends into the early twenties[4] (BOX 1). This transition to becoming an independent and responsible adult is inherently intertwined with adjustments in personal goals and motivations — for example, developing priorities related to career, identity, friends, romantic partners, family, community and religious or philosophical beliefs. This developmental transition involves greater use of cognitive control skills, such as the use of top-down effortful control to modify attention, emotion and behaviour in service of long-term 'adult' goals. However, social and affective processes also have crucial roles in these maturational changes[5,6]. An adolescent's success in pursuing long-term academic, athletic or artistic goals, for example, typically requires motivation to practice the relevant skills and a desire to persevere through difficulties, and these motivations are shaped by social experiences and are inherently intertwined with individual feelings about the value and relative priority of the goal.

There has been growing interest in understanding the neural changes that underpin these complex developmental processes. This has led to exciting scientific advances at this nexus of cognitive neuroscience, social neuroscience and developmental science. Investigations into these neuromaturational changes also hold promise for addressing some of the high-impact negative health problems that emerge in adolescence, including increased rates of accidents, alcohol and drug use, teenage pregnancies, depression and suicide, and violence[7–9].

Box 1 | **Adolescence from an anthropological perspective**

There is a commonly cited myth that adolescence was 'invented' by industrial society to extend occupational training beyond childhood. However, some of the neurobehavioural changes seen in human adolescence, such as increases in exploratory tendencies and changes in reward processing, have been observed in many non-human species as they go through puberty (BOX 4). Moreover, as documented by the anthropologists Schlegel and Barry[155] in a study of 186 pre-industrial societies, virtually every human society (including hunter-gatherers and pastoralists) recognizes an 'adolescent' period as a stage that is distinct from childhood but during which individuals are not yet fully adult in status. Thus, it is not the existence of adolescence as a developmental stage that has changed in recent history but rather the timing and length of this developmental period. That is, historically puberty occurred at relatively older ages (for example, age of menarche at 15–16 years of age) and taking on adult status typically ensued within 2–4 years. In contemporary society, puberty often occurs at much earlier ages (the mean age of menarche in the United States is 12 years and early signs of puberty typically begin by 9–11 years of age), whereas the process of achieving full adult roles is often stretched into the mid-twenties. Thus, in modern society, 'adolescence' has been stretched to span a much longer interval of development.

In addition, the social structures of adolescents have undergone major changes in recent human history, as have key aspects of developing long-term goals. In contemporary society, adolescents spend most of their time in school with same-age peers or in other structured educational and training environments, where the primary goal is to prepare the adolescent for occupations in a distant and abstract future. In pre-industrial societies, however, adolescence functioned primarily as a period of social and reproductive development[155] or apprenticing to learn directly utilitarian skills.

The relationships between these changes in the length, timing, nature and goals of adolescence and the brain changes associated with adolescence are not yet understood. For example, some aspects of adolescent development (for example, social–affective changes at puberty) occur at earlier ages, whereas other developmental milestones (for example, taking on adult roles and responsibilities in society) occur at later ages, raising the question how this differential timing of these external factors (combined with earlier activation of pubertal changes in social and affective processing) affects the development of neural systems that are involved in social and emotional regulation and the self-regulation necessary for taking on fully adult roles.

In this Review, we briefly discuss some of the prevailing views on adolescent brain development. Next, we review neuroimaging studies of cognitive control, affective processing and social processing in adolescence. In addition, we discuss the pronounced social and affective changes in adolescence, including the importance of interactions between cognitive, affective and social processing during this period of development[10,11]. Last, we suggest a re-evaluation and extension of the prevailing models of adolescent brain development. We emphasize the lack of data supporting any simple view of frontal cortical immaturity as the explanation for adolescent vulnerabilities, and consider the growing evidence for specific social–affective changes that begin during pubertal development as conferring increased vulnerabilities in some adolescent contexts. We also highlight the need for a better understanding of the neuromaturational underpinnings to these social–affective changes, including the role of pubertal development, and the potential value of investigating how these changes may contribute to unique opportunities for learning and adaptation in adolescence.

Current views of adolescent brain development

Over the past decade, our understanding of the neural mechanisms that underlie changes in cognitive, affective and social development during adolescence has increased tremendously. As will be reviewed, there has also been intense interest in applying this advancing knowledge to help inform broad societal issues, such as adolescent health, education and legal policies. Several influential models of adolescent brain development have proposed that a maturational gap between cognitive control and affective processes (including reward and threat processing) may explain adolescent increases in risks for engaging in impulsive and dangerous behaviour (for example, see REFS 2,7,8). These models tend to emphasize the relatively faster maturation of subcortical affective brain areas in comparison to more slowly maturing frontal cortical brain areas as the reason why adolescents tend to make more emotional (that is, less rational) decisions, resulting in actions that do not sufficiently weigh consideration of long-term outcomes.

Despite the appeal of these models in explaining the high rates of dangerous and impulsive behaviour in adolescents, it also is important to evaluate the degree to which the available neuroimaging data support these models. A number of research groups have begun to suggest that there has been too much emphasis on frontal cortical immaturity as the reason why adolescents engage in risky behaviour, and they have begun to point increasingly towards a more nuanced understanding of interactions across cognitive, affective and social processing[12]. There also is a growing recognition that social contexts strongly influence how these neural systems develop and how adolescents make decisions.

Neuroimaging adolescent development

Structural MRI (BOX 3) and functional MRI (fMRI) have been used to study how changes in brain structure and activity, respectively, are associated with changes in behaviour during development. In the past decade, a large number of fMRI studies have been conducted in the domains of cognitive, emotional and social development. In these studies, the typical age range of the subjects is 8–25 years, which provides a good framework for the examination of broad changes that occur during adolescence. However, as mentioned above, there is considerable variability in the ages used, many studies have gaps in the measurement of different phases of adolescence (for example, comparing only early adolescent 8–12-year-olds with adults or only comparing mid-adolescent 13–17-year-olds with adults) and most studies have only tested for linear age-related changes rather than testing for models of adolescent-specific patterns of change (for example, U-shape or inverted U-shape patterns of development). Furthermore, age-related changes provide a rough proxy for adolescent phases but do not permit examination of puberty-specific effects, and most of these studies did not include an assessment of pubertal development. Nonetheless, there is now an impressive set of fMRI studies through which to consider the developing brain and its role in adolescence-specific transitions in cognitive, affective and social processing and their interactions. Below we review and discuss these studies in the context of a meta-analysis (FIG. 1a; Supplementary information S1 (table)).

Box 2 | **Sex hormones in adolescence**

Pubertal development is associated with numerous changes in the brain, with evidence that hormone levels and neural function mutually influence each other. The single most important step in the onset of puberty occurs when the hypothalamus begins to release substantial amounts of gonadotrophin-releasing hormone (GnRH) in a pulsatile manner during sleep. This pulsing of GnRH begins the re-awakening of the hypothalamic–pituitary–gonadal (HPG) axis, which is first active during prenatal and early postnatal life (sometimes referred to as the neonatal 'mini-puberty') and then is shut down by inhibitory inputs to the hypothalamus, remaining quiescent throughout childhood. Pulses of GnRH stimulate the pituitary to produce the hormones follicle-stimulating hormone (FSH) and luteinizing hormone (LH), which in turn stimulate the ovaries and testis to produce the sex hormones oestrogen and testosterone, respectively. The mechanisms that trigger the re-awakening of pubertal GnRH pulsing are not fully understood, but they include interactions with neural systems involved in metabolic regulation, energy storage and sleep regulation. There has been rapid progress in understanding several aspects of the process over the past decade, including the importance of the hormone leptin (a protein manufactured in fat cells that has a key role in regulating energy intake, energy expenditure and appetite) and kisspeptins (a family of neuropeptides encoded by the *KISS1* (KiSS 1 metastasis-suppressor) gene that have been identified as the conduits for the effects of leptin actions on GnRH neurons in the hypothalamus).

A second neuroendocrine axis that forms a core aspect of pubertal maturation involves increases in growth hormone (GH) secretion from the pituitary, which has a crucial role in the rapid physical growth during this period. As with the gonadal hormones, this GH increase at puberty is also sleep-dependent. A third component of puberty involves increases in the secretion of a testosterone-like hormone from the adrenal glands called dehydroepiandrosterone (DHEA) — this process is the least well understood in terms of the neural systems that initiate and regulate it.

The main hormones that regulate the bodily changes and emergence of secondary sexual characteristics of puberty are testosterone, oestradiol and DHEA. The physical sex differences that emerge at puberty are in part attributable to differences in hormone levels (for example, higher oestradiol levels in girls and greater testosterone increases in boys) but also to differences in the distribution and types of hormone receptors in target tissues.

There is relatively limited knowledge of how these hormones influence adolescent brain development and specific behavioural, cognitive and affective changes during adolescence. Several research groups have begun to focus on the role of pubertal hormones on neurobehavioural changes in adolescence, with intriguing preliminary findings[3,83]. As discussed in those reviews[3,83], addressing these questions will require both conceptual and methodological advances. Animal experiments that examine neural and behavioural changes associated with specific aspects of pubertal maturation and clinical studies that examine neural changes in response to hormone treatments (for example, the administration of oestrogen to pre-adolescent girls with Turner syndrome[156]) can provide additional insights.

Functional MRI studies of cognitive control

It is well recognized that during adolescence, there is a steady increase in the ability to use cognitive control over thoughts and actions[13,14]. Cognitive control abilities start to emerge in early childhood and gradually improve over childhood and through adolescence[15,16]. These abilities and are often seen as a driving force behind cognitive development[17], and these increases in cognitive control abilities in adolescence mark a period of significant advancements in learning and successful adaptations to a wide variety of social contexts and cultural influences. For example, the ability to exert cognitive control over thoughts and actions is of crucial importance to success in most classroom settings — not only for the direct learning of skills such as reading, maths and the capacity to reason about abstract ideas but also at the level of behavioural control that supports sitting at a desk, avoiding distractions and doing homework.

Many developmental fMRI studies have been conducted in this domain, including investigations of basic cognitive control functions and more complex cognitive control functions in which different basic functions have to be combined. Although these functions are separable in their contributions to complex behaviour[18], they rely on overlapping areas in the lateral prefrontal cortex (PFC) and parietal cortex (also see REF. 19). However, the extent to which these brain areas are activated across development differs between studies and samples, as discussed below (FIG. 1a).

Basic cognitive control functions. Many studies of basic cognitive control functions, such as working memory, inhibition and interference, and task switching, have reported that regions involved in these functions in adults (including the lateral PFC and parietal cortex) become increasingly engaged during childhood and adolescence (FIG. 1a; Supplementary information S1 (table)). For example, in spatial and verbal working memory paradigms that contrast high working memory load with low working memory load, increases in activity in the ventral and dorsolateral PFC and parietal cortex have been reported when 7–12-year-olds were compared with adults; when 7–12-year-olds were compared with mid-adolescents (13–17 years) and adults; and for linear comparisons from the age of 7 years to adulthood[20–30]. Studies using response inhibition or interference suppression tasks report an age-related increase in activation in the inferior and middle frontal gyrus in 'go' versus 'no-go' trials when children and early adolescents (7–12 years) were compared with adults; when children and early adolescents (6–12 years) were compared with mid-adolescents (13–17 years) and adults; and for linear comparisons from the age of 7 years to adulthood[31–36]. In addition, several task switching studies have reported increased activation in the lateral PFC and parietal cortex in adults relative to children (ages 7–12 years) and adults versus mid-adolescents (ages 10–17 years or 13–18 years) in 'switch' versus 'repeat' trials[37–39]. These findings have been interpreted as indicating that areas of the PFC have a slow developmental trajectory and are

Box 3 | **Structural brain development in adolescence**

Numerous structural neuroimaging studies have demonstrated that adolescent development involves widespread changes in the brain. Longitudinal research examining changes in brain structure over time has shown that cortical white matter throughout the brain increases with age throughout childhood and adolescence. By contrast, cortical grey matter, which reflects neuronal density and the number of connections between neurons, follows an inverted-U shape over development, peaking at different ages depending on the region[157–159]. Within the cortex, grey matter reduction is most protracted for the dorsolateral prefrontal cortex and the temporoparietal junction; here, cortical grey matter loss continues until the early twenties[159,160]. The development of subcortical brain regions is also subject to both linear and nonlinear changes, such that some subcortical regions (such as the caudate and the putamen) linearly decrease in size throughout adolescence, whereas other subcortical regions (such as the amygdala and the hippocampus) show an increase in size at the onset of puberty, after which growth stabilizes in adolescence and adulthood[161]. These dynamics of structural brain development have been summarized in several excellent reviews (for example, see REF. 160).

not engaged to the same extent in children and adolescents as in adults[40].

However, many studies of the same basic cognitive control functions have found age-related decreases in frontal cortical activity in early adolescents compared with children and adults, mainly in the superior part of the lateral and medial PFC (FIG. 1a; Supplementary information S1 (table)). These decreases were found for different domains of working memory (in studies comparing ages 6–12 years, 13–17 years and adults)[41–43], for response inhibition (in studies comparing ages 6–12 years versus adults)[32,44–48] and in task switching (in studies comparing ages 8–13 years versus adults)[49,50]. These findings have been interpreted as indicating increased efficiency of these networks over time. However, it is difficult to confirm this interpretation because these decreases in activation are not always accompanied by performance differences.

Thus, although the parietal cortex seems to show a relatively consistent pattern of increased activation in cognitive control tasks with increasing age (except for one study that showed a decrease with increasing age[22]), studies of lateral and medial PFC show both increases and decreases in activation, depending on the task paradigm and the PFC subregion involved in the task (FIG. 1a). Moreover, mid-adolescence-specific increases (for ages 13–17 years relative to both 6–12 years and adults) have been reported for regions in the lateral PFC in working memory, inhibition and task switching tasks[39,41,47,51,52] (FIG. 1a; Supplementary information S1 (table)); such an inverted U-shaped relationship between age and activation could be due to increases in task engagement in adolescents compared to children and adults.

Taken together, it is difficult to reconcile how this degree of variability in neuroimaging findings in the development of basic cognitive control functions provides support for the model of 'frontal cortical immaturity' or the concept of 'linear advances in PFC development' across adolescence. Indeed, if such varied findings of increases or decreases in activation can be interpreted as consistent with the concept of frontal cortical immaturity, this would seem to render the model as virtually unfalsifiable. Our meta-analysis suggests that such a simple model of increased activation in the PFC is unlikely to account for the developmental transitions in basic cognitive control that take place during adolescence[53,54].

Complex cognitive control functions. Several recent studies have used approaches that involve more complex cognitive control tasks, such as performance monitoring, feedback learning and relational reasoning, which require a combination of basic cognitive control functions[18]. This approach can detect strategy differences between people in a particular task. These studies have revealed interesting developmental trajectories of PFC activation (Supplementary information S1 (table)). For example, performance monitoring studies (that is, studies involving error and feedback processing) that included early adolescent (ages 8–12 years), mid-adolescent (ages 13–17 years) and adult age groups did not confirm the strict frontal cortical immaturity view[55–59]. Instead, these studies report that the frontal cortical network was engaged to the same extent in participants of different age groups but under different experimental conditions. Specifically, in early adolescents, the PFC and parietal cortex were activated following positive performance feedback, whereas in adults, the same regions were activated to the same extent following negative feedback, with mid-adolescents showing a transition phase[57,58]. A similar nonlinear pattern was found in a relational reasoning task[60] in early adolescents, mid-adolescents and young adults (ages 11–30 years). When subjects were asked to combine and integrate different spatial dimensions (that is, relational integration), only mid-adolescents (14–18 years) showed increased activation in the anterior PFC. The authors interpreted this as reflecting a mid-adolescence-specific cognitive strategy to perform the task in an efficient way (see REFS 61–63 for other examples of relational reasoning studies). Indeed, the increased activation in mid-adolescence was associated with faster reaction times and increased accuracy. However, the exact relation between neural activation, task performance and strategy use is not well understood at this time.

Flexibility for recruiting cognitive control systems? The question then arises: what is the general pattern that emerges from fMRI studies on cognitive control? The data discussed so far provide evidence against the view that these brain regions simply come increasingly 'online' with advancing age through adolescence. Instead, the high degree of variability in the findings could reflect a less automatic and more flexible cognitive control system in adolescence. It is possible that the degree to which cognitive control processes are engaged or activated in adolescence are strongly influenced by the motivational salience of the context. Factors such as the presence of peers, task instructions, strategies and the affective appraisal of the value or priority of performing the task may have relatively large influences on the extent to which cognitive control systems are recruited in adolescence[53]. As will be discussed in later sections, there is growing recognition that social and affective factors are particularly important in influencing aspects of adolescent engagement. The ability to quickly shift priorities,

Relational reasoning
An essential component of fluid intelligence that requires a number of verbal or spatial dimensions to be considered simultaneously to reach a correct solution.

adjusting the degree of cortical activation in a given task or situation according to the social and motivational context could contribute to greater variability in cognitive control. However, this flexibility in making quick adjustments in the degree of engagement across changing contexts may be crucial to the ability of youths to learn about and adapt to rapidly changing adolescent social contexts. For example, adolescents are often the fast-adopters of social change — such as learning new trends in language, technology, music and fashion or when adapting to new cultures after immigration[64].

Interestingly, two longitudinal studies that followed adolescents (ages 8–23 years and 15–18 years) over a period of 3 years reported no age-related changes in activation in the frontoparietal network in a feedback-learning and working memory paradigm. Instead, changes in task performance in the same person measured at different ages correlated with changes in activation in the lateral PFC[65,66]. Furthermore, in a working memory training study, young adolescents (ages 11–13 years) showed increased activation in the lateral PFC after 6 weeks of practice, whereas before training these adolescents showed less activation in the lateral PFC compared to adults before training[67]. These findings support the idea that frontal cortical brain regions in adolescence are sensitive to context and that their activity can be enhanced by training. The training study[67] indicates that this flexibility of the frontal cortical network may be greater in adolescence than in adulthood, although further studies are needed to confirm these findings.

This proposed flexibility in frontal cortical networks in adolescence is further supported by studies on functional connectivity in the absence of a behavioural task (that is, resting state analyses). These studies have demonstrated that the main circuitry for cognitive control is already in place at the start of adolescence[68], but the strength of connectivity within this circuitry continues to undergo maturational changes across adolescence. For example, there is a tendency for short-range connections to become weaker with age, whereas long-range connections, which are important for integration across circuits, become stronger with age[69]. The authors of this study interpreted their findings as consistent with a model of developing tighter 'integration' of some regions into long-range networks over time, while segregating the short-range connections between other sets of regions into separate networks. Because the long-range connectivity patterns are still undergoing maturational strengthening, it is likely that some aspects of integrative cognitive control may be less automatic and more flexible during adolescence. As a result of weaker connectivity across these long-range integrating circuits, adolescents may be more vulnerable to variability in performance under high demands on attentional and decision-making networks in some situations (because the ability to integrate control is less automatic); however, these same qualities (less automatic responses) may also enable adolescents to respond in creative and adaptive ways. Most importantly, however, these findings suggest that adolescence is a crucial time of development during which specific learning (or training) experiences may actively sculpt final connectivity patterns in some of these long-range cognitive control networks (see REF. 70 for a training study on functional connectivity in adults supporting the view of adaptive change in the frontal cortical circuitry).

Taken together, findings from the few existing longitudinal and training studies (which are more powerful in detecting the trajectories of brain change than the more usual cross-sectional studies comparing individuals of different ages) highlight the complexities of disentangling specific developmental changes during adolescence. An important goal for future research will be to parse the developmental changes in brain activation that reflect four relatively different processes that could influence cognitive performance in adolescents: maturational changes in the fundamental capacity to perform a task; other task-relevant factors, such as degree of engagement, motivation and sensitivity to social and affective context; changes that reflect the direct effects of training; and developmental changes in the capacity for learning and training. As reviewed below, it appears that task performance (and perhaps some developmental learning effects) in adolescents may be particularly sensitive to social and affective influences.

Functional MRI studies of affective processing

Neural systems that underpin affective processing can be conceptualized not only as systems involved in emotions and motivation but also, more broadly, as a network of 'valuing' systems that are involved in learning about rewards and threats and in regulating 'approach' and 'avoidance' behaviours accordingly. During adolescent development, the most salient types of rewards and threats typically reside in the social domain (for example, being admired, accepted or rejected by peers and early romantic and sexual experiences). Accordingly, it is important to recognize the inherent overlap between affective and social processing in adolescence. However, to date, most studies in this area have focused on monetary rewards to examine how the ventral striatum, which is a subcortical brain region that is active when a person receives or expects a reward[71], responds to risks and rewards in adolescents compared to adults (FIG. 1b; Supplementary information S1 (table)).

Box 4 | **Animal research on puberty-specific changes in reward processing**

There is compelling evidence from animal models showing that changes in gonadal hormone levels in puberty induce a (second) organizational period to guide the remodelling of the adolescent brain in sex-specific ways[162,163]. Rodent studies have also shown a remodelling of the dopaminergic systems involved in reward and incentive processing in the peri-adolescent period. This remodelling involves an initial rise in dopamine receptor density, starting in pre-adolescence, and a subsequent reduction of dopamine receptor density in the striatum and prefrontal cortex[163] — a pattern that is more pronounced in males than females[163]. As a result, dopaminergic activity increases substantially in early adolescence and is higher during this period than earlier or later in development[163]. The developmental changes in reward processing in animals in these studies are similar to those emerging from the human functional MRI literature. Given the important role of dopamine in reward processing, the developmental changes in dopamine receptor levels may be linked to the increase in novelty seeking, exploratory behaviour and reward-seeking behaviour at puberty[164,165]. Thus, translational research focusing on the mechanisms that underpin pubertal changes in reward responses may provide important insights into human adolescent behaviour.

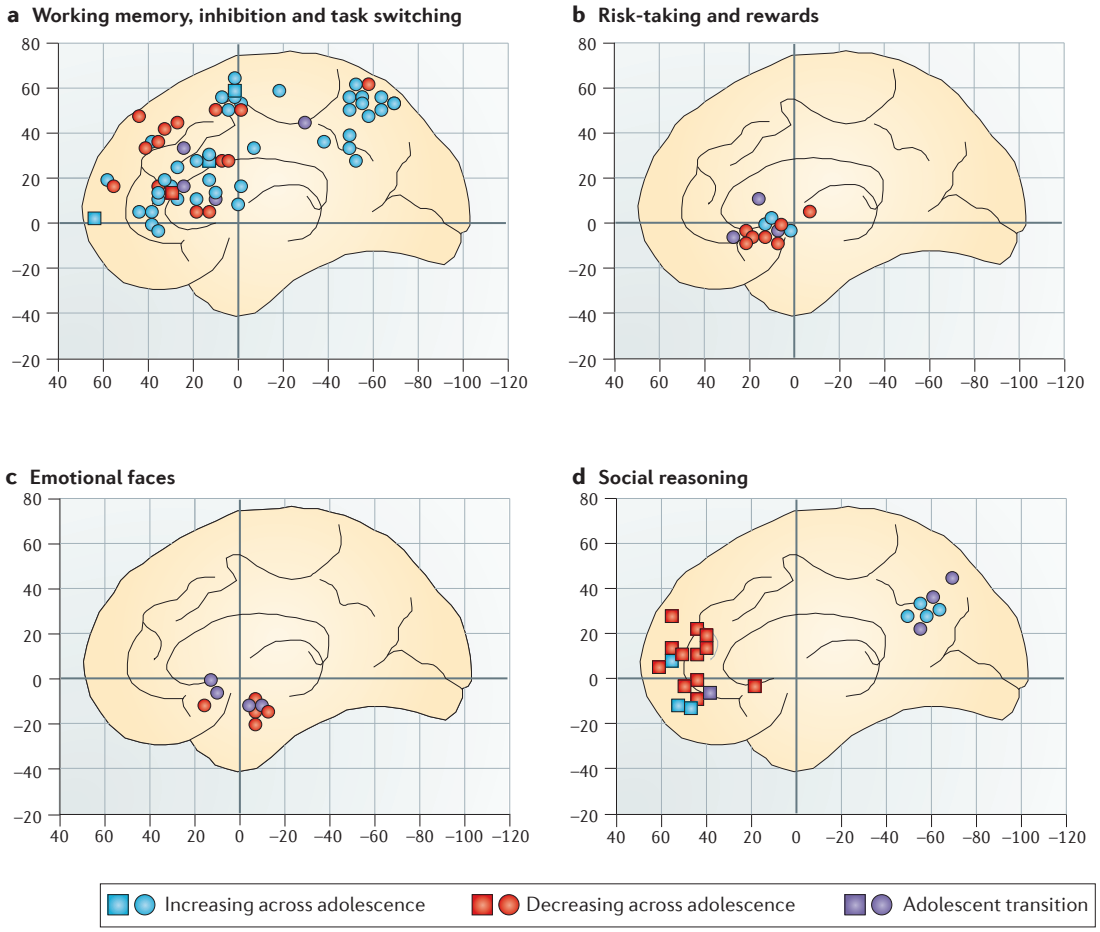

Figure 1 | **Meta-analysis of functional MRI studies in adolescents.** Results from a meta-analysis of a representative set of functional MRI studies, which were conducted between 2001 and 2011, of cognitive, affective and social processing in adolescents compared to other age groups. **a** | Frontoparietal and anterior cingulate cortex activation in working memory[20–30,41–43,52,65,167], inhibition[31–36,44–48,168] and interference suppression and task switching studies[35,37–39,49,50,78]. **b** | Striatum activation in reward processing studies[72–82,169]. **c** | Amygdala and striatum activation for face processing studies[89–94,96,100,170]. **d** | Anterior medial prefrontal cortex and temporoparietal junction activation in social–cognitive reasoning studies[107–113,115,116,127,128,130,171,172]. For illustrative purposes and for reasons of comparability, a slice of the mid-brain is shown ((x = 0), Montreal Neurological Institute coordinates) but the activations are displayed as circles when activation was in ventral and dorsolateral prefrontal cortex (x > +/– 20) or in superior frontal sulcus–frontal eye fields and parietal cortex (x > +/– 20), and as squares when activation was in the medial prefrontal cortex (x < +/– 20). The findings of the reviewed studies are summarized as 'increasing across adolescence' (light blue squares and circles), which indicates that the specific region is more engaged with increasing age; 'decreasing across adolescence' (red squares and circles), which indicates that the specific region is less engaged with increasing age; and 'adolescent transition' (purple squares and circles), which indicates that mid-adolescents process information differently from both children and adults. It should be noted that: first, the increases and decreases were dependent on the contrast used and therefore should be interpreted in this context (see REF. 173) and, second, not all studies used more than two age groups — a design that does not allow for an examination of transitions. Supplementary information S1 (table) provides an overview of all studies that were included in the meta-analysis, including the age range and sample size for each age group.

When receiving rewards, adolescents (ages 12–17 years) consistently show increased striatal activation relative to children (ages 7–12 years) and adults[72–80]. By contrast, adolescents tend to show less activation in the striatum than adults during reward expectation or anticipation (that is, when participants observe a cue that indicates a potential reward)[76,81,82]. The differential response to cues (reward anticipation) and actual receipt of rewards in adolescents may help to explain some of the inconsistencies with regard to ventral striatum activity in adolescents[83]. For example, the finding that underactivation (or no change) in ventral striatum activity is found during reward anticipation in adolescents may help to explain why some studies show no differences between adolescents and adults in risk-taking behaviour, despite pronounced neural differences during reward processing[73,84–86].

One way in which reward processing may influence decision-making is through the prediction error. Reward prediction error signals reflect the difference between the expected value of an action and the actual outcome of the action, and are encoded by phasic

activity in the mesolimbic dopamine system (including the ventral striatum). These prediction error signals appear to have a crucial role in the process of learning and adjusting behaviour to adapt to changing contexts or conditions. The first developmental study[56] of prediction error signals in children, adolescents and adults found that prediction error signals in the striatum were highest in adolescents, whereas decision-value signals in the medial PFC did not show a consistent developmental pattern. Results of a second developmental study of reinforcement learning did not implicate the prediction error signal directly but pointed to the connectivity between the ventral striatum and medial PFC as the source of developmental differences in how learning signals guide adolescent behaviour[87]. Interestingly, recent evidence has demonstrated that value-based decision processes are based on neural computations that use the subjective value of the expected reward[88], again implicating interactions between reward prediction at the level of the ventral striatum and higher-level, cortical processing of 'valuing', which is likely to incorporate more subjective aspects of valuing, such as the social or affective context. Taken together, these findings point to a promising line of investigation into the mechanisms by which subcortical value-based inputs may interact with cortical value-based inputs to signal motivational salience.

In addition to these studies of reward processing, a number of investigations have examined developmental changes in the response to threat stimuli. For example, increased activity in subcortical brain regions has been observed in adolescents in response to emotional faces (FIG. 1c). Several studies have reported enhanced activity in the amygdala, a region of the brain that is important for the processing of negative affect, in mid-adolescents (ages 12–18 years) compared with adults when looking at pictures of fearful faces[89–94] (see REFS 95–99 for studies that focused on other brain regions or younger children). Pictures displaying positive emotional (for example, happy) faces induced more activation in adolescents relative to adults in the ventral striatum — the area that is also more active in response to receiving rewards in mid-adolescents relative to adults[94,100]. Thus, it appears that mid-adolescence is associated with a more general intensification of affective processing, not only in the approach — or positive affect — domains (such as rewards and happy faces) but also for stimuli that may signal threat and avoidance (that is, fearful faces).

Together, these findings suggest that the neurodevelopmental changes in affective processing in approach and avoidance follow nonlinear developmental patterns, with a peak in subcortical brain activation in mid-adolescence. This pattern may underlie part of the intensification of emotional and motivational experiences in mid-adolescence, and this intensification of affect may create new challenges to emotional regulation and self-control[101]. Moreover, the increased activity in 'valuing' systems in adolescence may reflect a sensitive period for learning about sources of reward and threat, particularly in social domains.

Social–cognitive development
Changes in cognitive skills and knowledge that facilitate understanding social situations, such as mentalizing and perspective-taking abilities.

Social–affective development
Changes in motivational and emotional aspects of social processing (such as empathy, increases in the salience of obtaining status, admiration and affiliation from peers) and the development of affective skills that support social competence.

Mentalizing
The ability to infer mental states of others, such as one's intentions, beliefs and desires — a key dimension of social–cognitive development in adolescence.

Functional MRI studies of social development

The fundamental maturational task of adolescence is achieving adult social competence — that is, developing the knowledge and skills to be capable of functioning independently from parents or other responsible adults. Adolescents appear to be naturally motivated to want greater independence from their parents and to establish their individuality[102]. Adolescents are drawn to build and explore new social networks (that is, peer groups) and to increase prioritization around peer issues of belonging, acceptance and interests in romantic and sexual partners. Achieving success in these domains requires new social skills, social knowledge, affect regulation, adaptive coping skills and, in general, improved social competence[103].

There has been recent progress in understanding neural systems relevant to two dimensions of social development in adolescence: social–cognitive development, which concerns the knowledge and capacity to understand social situations, and social–affective development, which concerns the motivational and emotional aspects of social skills.

Social–cognitive development. There has been considerable progress in understanding the development of neural systems that underlie social–cognitive skills such as mentalizing[104]. Basic social detection and theory-of-mind develop in early childhood, whereas more complex social–cognitive skills, such as mentalizing and metacognition, mainly develop in adolescence. The development of complex social–cognitive skills is probably driven partly by environmental demands and experiences, such as the greater need to adapt to the peer group and newly emerging romantic interests. Such social–cognitive skills become increasingly important as adolescents learn to adapt to rapidly changing social environments, in which the opinions and evaluations of peers become increasingly salient.

Recently, researchers have identified a 'social brain network' — a network of brain regions, including the medial PFC and temporoparietal junction (TPJ) — that is important for mentalizing and perspective-taking[105] and that undergoes structural and functional changes during development[106]. Studies using mentalizing and social interaction paradigms have shown that specific regions in the social brain network contribute to the development of intention understanding in social reasoning in children and adolescents (see REF. 106 for a review). As highlighted in our meta-analysis (FIG. 1d), studies using social reasoning paradigms[107–114] and self-knowledge paradigms[115,116] have shown that the medial PFC is often more activated in adolescents (ages 9–18 years) compared to adults[106], whereas the TPJ is often less activated in adolescents (ages 10–17 years) compared to adults[106].

One of the main changes in the nature of social interactions in adolescence is the shift from self-oriented behaviour towards other-oriented (that is, pro-social) behaviour[117]. These changes enable the formation of more complex social relationships and are particularly important for functioning in peer groups — adolescents have a stronger motivation for peer acceptance compared with children and adults[118]. Social interaction paradigms

can be used to investigate neural activity associated with self-oriented thoughts and other-oriented thoughts and actions. Inspired by classic social utility models of decision-making, social psychologists have developed experimental 'games' in which two-person interactions are investigated in a laboratory setting. According to social utility models, social behaviour is generally motivated by self-gain and by concern for others[105]. The latter is essential for other-oriented behaviour and requires the ability to consider other people's feelings, thoughts, intentions and actions, therefore drawing heavily on theory-of-mind (that is, perspective-taking) abilities. Comparison of self-gain and other-gain is involved in social judgements of fairness and reciprocity, which in turn have important roles in the display of other-oriented behaviour. Therefore, these games provide a valid experimental context for studying these important aspects in the development of self- and other-oriented processes[105].

Two of the most commonly used games to study social decision-making in adults are the Ultimatum Game[119] (FIG. 2) and the Trust Game[120]. These games have proven to be highly useful for studying developmental differences in self- versus other-oriented thoughts[105]. Studies using these games have found that self-oriented thoughts decrease and other-oriented thoughts increase with age, with a transition phase around mid-adolescence (ages 12–16 years) during which other-oriented thoughts become more dominant than self-oriented thoughts. In addition, these studies showed that children and early adolescents (ages 9–12 years) have less understanding of other people's intentions when making or judging decisions and, with age, increasingly take the perspective of others into account[121–123]. A meta-analysis demonstrated that these games activate brain regions that are implicated in the different value computations of social interaction, such as the valuing of self-gain versus gains for others[124]. That is, the brain regions that are involved in social cognition (anterior medial PFC, TPJ and insula) are involved in judging fairness and in reciprocating trust, and activity in these regions depends on perspective-taking demands[125,126].

Age comparisons using these games have demonstrated that with increasing age, adolescents are increasingly responsive to the perspective of another player. Concurrent with this behavioural change, there was a gradual increase in activation in the TPJ (and the dorsolateral PFC) and a gradual decrease in activation in the anterior medial PFC across adolescence[123,127,128]. The increase in TPJ activation correlated with the perspective-taking behaviour, independently of age, confirming the role of this area in perspective-taking[128]. The overactivation in the anterior medial PFC and underactivation in the TPJ in adolescents relative to adults mentioned above could be interpreted as underlying the decrease in self-oriented thoughts and actions and the increase in other-oriented thoughts and actions, respectively, that occur across adolescent development.

It is important to recognize that some of these developmental changes in fairness and reciprocity appear to reflect changes in explicit social knowledge and understanding; however, some of these changes may involve implicit learning processes and rely on the development of social–affective skills. Indeed, considerations of self and other's outcomes appear to be influenced by the social environment of adolescents. For example, there is evidence that popular adolescents (that is, those frequently liked and seldom disliked by peers) generally help, share and cooperate with peers and score highly on measures of empathy and perspective-taking[129].

Social–affective development. There is growing understanding of the neural systems that underlie aspects of social–affective development in adolescence. For example, studies on empathy[130] and social acceptance and rejection[131–134] have reported differences in brain activity between children, adolescents and adults in brain areas involved in processing affect and social pain, including the temporal pole and the insula (Supplementary information S1 (table)).

One study[131] that examined, in different age groups (ages 8-10 years, 12–14 years, 16–17 years and adults), neural activation in response to social acceptance and rejection from peers found increased activation in the ventral anterior cingulate cortex (ACC) and striatum in each age group when a participant received feedback that a peer liked them compared to feedback indicating that a peer did not like them. This is consistent with the idea that social acceptance is salient across these age groups and continues to be salient in adulthood. Social rejection was associated with activation of the insula and dorsal ACC in all age groups, but only adults showed additional recruitment of the dorsolateral PFC, which may indicate a better capacity to regulate rejection, although this was not tested using behavioural measures. In a study using the Cyberball game to elicit feelings of rejection, early adolescents (ages 10–12 years) showed more activation in the subgenual ACC during rejection than adults[135]. Activity in this region was associated with greater rejection-related distress in youths in a different Cyberball study[136]. Activity in the insula (which was also associated with greater rejection-related distress)[136] was reduced in individuals who have many friends in daily life (in the 2 years before the fMRI scan)[137], suggesting that young adolescents who had developed strong friendship networks were less sensitive to social rejection. Finally, this same research group also showed that increased subgenual ACC and medial PFC activity to social exclusion in the 12–13-year-olds predicted increased depressive symptoms in the year following the Cyberball study[138].

Taken together, these findings show promising approaches to investigating the development of social–affective processing in adolescence; however, they also raise a number of questions. One particularly thorny set of issues focuses on questions about the direction of effects. For example, some changes in neural activation in response to social and affective stimuli may depend on new patterns of social learning and experience in adolescence (such as greater reaction to social rejection secondary to affective learning that is simply more likely to occur during this period of development). By contrast, changes in the neural systems that underpin the motivational

Self-oriented thoughts
Concern for outcomes that benefit one's own gains, such as in economic exchange when benefits for self and benefits for others are often conflicting.

Other-oriented thoughts
Concern for outcomes that benefit others, even when this is at the expense of gains for self, such as when evaluating what is fair for two parties.

Trust Game
Two-person interaction game that requires perspective-taking and relies on feelings of fairness and concern for others.

salience of peer rejection may undergo maturational changes that render the systems biologically more reactive. It also seems likely that bidirectional effects could occur (maturational changes that fundamentally alter the motivational salience or reactivity that also interact with learning experiences that are more likely to occur in adolescence). Such bidirectional interactions could contribute to spiralling effects over time, such as sensitivity to rejection and a pattern of negative experiences leading to the development of depression in adolescence. Studies of high-risk and clinical samples followed over time will be needed to test these hypotheses.

There is also a need to focus on the specific role of puberty as a neurodevelopmental mechanism that may contribute to the increase in motivational salience of social learning relevant to depression. For example, there is evidence that the increased risk for depression in adolescence is linked to the increase in gonadal hormone levels[139]. Given the finding that neural activity during social rejection at ages 12–13 years predicted later depression, this suggests that pubertal hormones may influence social–affective development, perhaps by increasing the affective salience (and vulnerability to long-term consequences) of social rejection.

Puberty and social–affective changes

There is growing evidence that some of the social and affective changes that occur in adolescence are linked

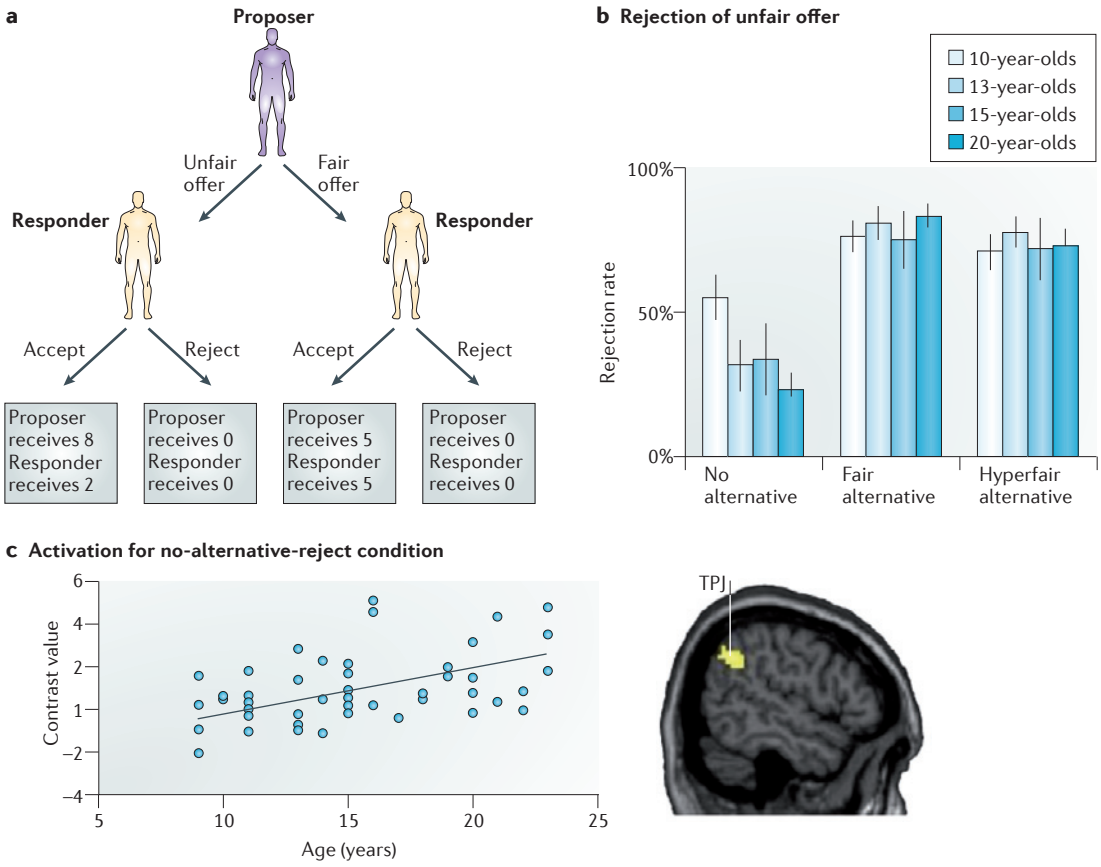

Figure 2 | **Interactive decision-making paradigms to examine social reasoning.** **a** | An example of the Ultimatum Game (UG) — a two-person interaction game that requires perspective-taking and relies on feelings of fairness. The game involves a proposer and a responder. The proposer can divide a fixed amount of money between the two players, and the responder decides whether to accept or reject the offer. When the offer is accepted, both players receive the stake according to the offer. When the responder rejects the offer, both players receive nothing. **b** | To vary perspective-taking demands on the responder, studies have made use of the mini-UG, in which the proposer is given two money-dividing options by the computer. One option is always an unfair division (8 for proposer, 2 for responder), and depending on the experimental condition, the second option can be unfair as well ('no alternative for proposer condition'), a fair split ('fair alternative condition') or a split that gives the advantage to the responder ('hyperfair alternative condition'). Results from behavioural tests show that in the mini-UG, responders take into consideration the options that the proposer had[121,127]. That is, unfair offers are mostly rejected when the alternative was fair or hyperfair but are more often accepted when the alternative was also unfair (in other words, the proposer could not help it but was restricted by the offers from the computer). Developmental studies have shown that in the no-alternative condition, which relies most on perspective-taking skills of the responder, there was an age-related decrease in rejection, indicating that the ability to understand the perspective of the first player increases with age. **c** | This increase was accompanied by increased activation in the temporoparietal junction (TPJ)[127].

to the onset of puberty[1–3,83,140]. Studies have focused on the role of the onset of puberty in the social re-orientation towards peers[10], in changes in neural processing of reward[141] and as shifting the balance of affective processing (with relatively more reward versus threat processing) interacting with cognitive control[2,101] in adolescents. Despite considerable evidence that puberty is linked to the increases in sensation-seeking and some aspects of risk-taking that occur in adolescence[3,9,83], there is little understanding of the specific hormonal changes that influence the development of those neural systems involved in motivational or emotional tendencies towards sensation-seeking behaviour. More generally, relatively few studies have investigated the role of puberty versus the role of age per se or the role of specific hormones in these behavioural changes.

We believe there are several reasons why it is important to investigate the role of hormonal changes in puberty at the interface between social and affective processing. For example, there is growing evidence that increases in risk-taking in adolescence emerge after the increase in sensation-seeking associated with puberty and occur primarily in affective salient social contexts. That is, adolescents show greater risk-taking than adults or children primarily when they are with peers (or believe they are being observed by peers), and such 'peer' effects are evident in both real-life and laboratory studies of risk-taking[77,142]. Greater risk-taking in adolescence has also been reported in emotionally charged (or 'hot') situations, but no adolescent increases in risk-taking occur in low-affect (or 'cool') contexts in the same experimental task[143]. On the basis of these and other findings (as discussed below), we propose that changes in gonadal hormone levels at puberty contribute to adolescent risk-taking through two interacting effects, namely by increasing the motivational salience of acquiring social status and by increasing the tendency to seek novel and high-intensity affective experiences — particularly in social contexts that create opportunities to gain peer admiration.

Moving forward: new heuristic models

On the basis of the findings reviewed above, we highlight what we regard as two important challenges facing the field regarding the prevailing models of adolescent brain development. First, the prevailing models are typically used to address broad issues of clinical relevance and social policy in ways that emphasize frontal cortical immaturity (or a maturational 'gap' in cognitive control) to explain the emergence of risky, impulsive and dangerous behaviours in adolescents. As described above, neuroimaging studies in adolescents do not support these aspects of the prevailing models. Rather, the data point to an adolescent flexibility in cognitive engagement, depending on the social and motivational context. The exciting challenge is to better understand how these incentives exert such strong influences on adolescents' engagement, decisions and behaviour — not only in ways that create vulnerabilities towards unhealthy incentives but also in ways that create unique opportunities for learning, adaptation and positive motivations relevant to health, education and social development in adolescence.

Second, the prevailing models are based on cognitive neuroscience studies that have relied primarily on cross-sectional comparisons between samples of 'adolescents' and 'children' and/or 'adults', and these groups have typically been defined by widely varying age ranges across different studies and laboratories. As a result, the current understanding of the maturational processes that underlie adolescent development is limited. One important example is the need to better understand the role of pubertal maturation on specific neurodevelopmental processes. We believe that this challenge will entail addressing not only methodological issues (for example, conducting studies designed to disentangle age and pubertal effects) but also conceptual issues (for example, refining models to address the role of specific hormones on specific aspects of social and affective development).

Below, we offer suggestions on how these challenges can be tackled and present a model of adolescent brain development that includes a focus on the role of puberty (FIG. 3). Our model proposes that the combination of flexibility in PFC recruitment and changes in social–affective processing can create vulnerabilities to engaging in negative behaviours in some incentive situations but is generally adaptive and developmentally appropriate to the tasks and learning demands of adolescence. There are two key aspects to this model. The first focuses on social–affective engagement and goal flexibility; and the second focuses on the role of pubertal hormones in social–affective engagement.

Social–affective engagement and goal flexibility. As described above, there is growing evidence that adolescence is a developmental period during which the degree of cognitive engagement is relatively flexible, depending on the social and motivational salience of a goal. This flexibility (and sensitivity to social and affective influences) may confer greater vulnerabilities for adolescents to act in ways that appear impulsive and immature, such as placing greater motivational value on gaining peer admiration for a daring action than considering the risks and long-term health consequences of that behaviour. However, this capacity to quickly shift goal priorities may also enable adolescents to effectively engage cognitive systems in situations in which they are highly motivated to do so and in ways that facilitate learning, problem-solving and the use of divergent creative abilities[144]. Indeed, emerging evidence from animal studies supports the idea that juveniles can outperform adults in some complex cognitive tasks (BOX 5).

Our model also is consistent with the idea that adolescence is an important period for developing cognitive control skills through training and experience. When adolescents are motivated, their capacity to engage can result in quick mastery of complex tasks. Consider, for example, a tedious and precision-demanding task such as using cell phone text messaging to communicate with peers — individuals who have learned these skills in adolescence typically reach a higher level of mastery than those who have learned as adults.

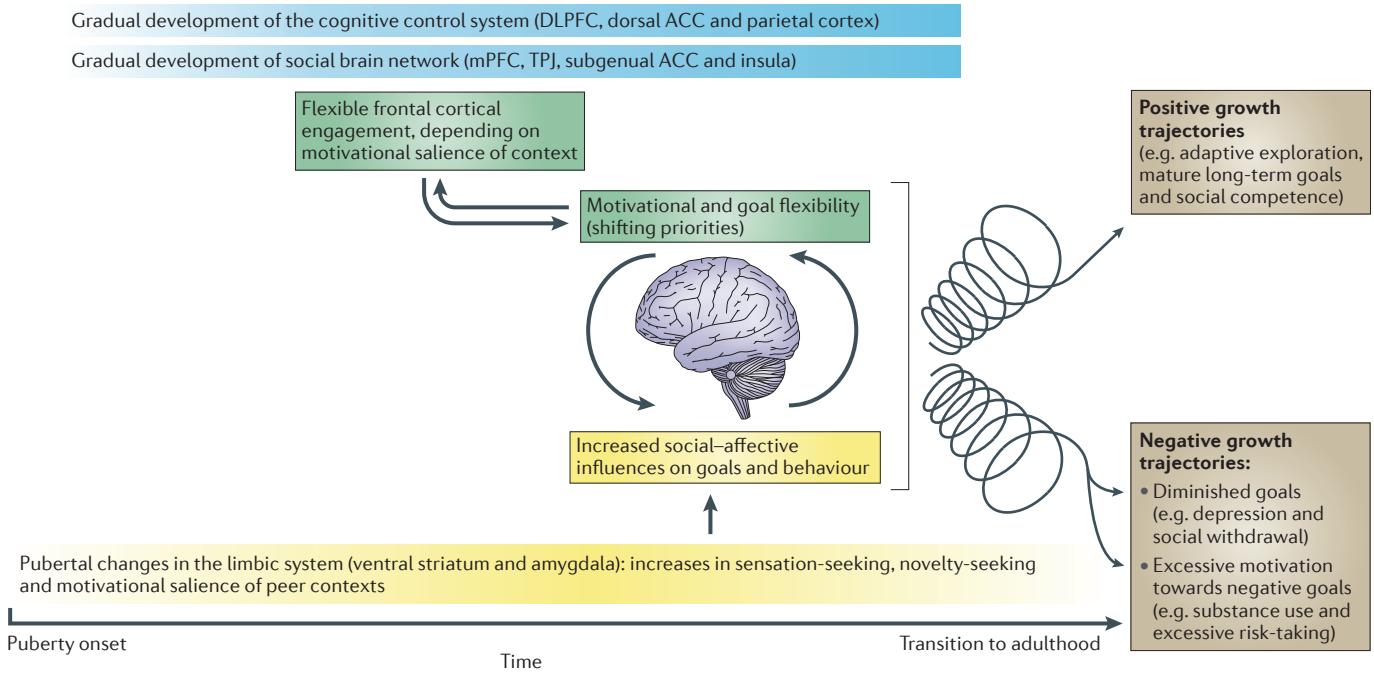

Figure 3 | **A model of adolescent brain development.** This figure illustrates a proposed model of adolescent brain development that begins with changes in social and affective processing (yellow boxes) associated with the onset of puberty. Specifically, rapid increases in hormone levels at the onset of puberty influence the development of limbic circuits, probably by inducing changes in the ventral striatum and amygdala (these regions have a pre-eminent role within the broader corticostriatal circuitry, which enables affect-laden stimuli to influence goals and behaviour). These pubertal changes contribute to increases in novelty-seeking, sensation-seeking and a tendency to process status-relevant social stimuli (for example, receiving attention and admiration from peers) as having increased motivational salience. Although these social and affective changes begin early (near the onset of puberty), they appear to peak in mid-adolescence and continue to influence behaviour, decisions and learning throughout several years of adolescent experiences (indicated by the colour gradient in the bottom yellow box). These social and affective influences interact with a broader set of changes in cognitive control and social cognitive development (blue boxes), which includes the acquisition of social and cognitive control skills that develop gradually across adolescence. These interactions between social–affective processing systems and cognitive control systems contribute to flexibility in the engagement of frontal cortical systems in adolescents, depending on the motivational salience of the context. In many contexts, these changes lead to increased social motivation and tendencies to explore, take risks and try new things — particularly when such bold behaviours may bring admiration from peers. An important feature of this model is the prediction that this increase in social–affective engagement not only influences incentives and behaviour in the moment (for example, choosing a specific bold but risky action to impress peers) but also influences motivational learning and patterns of behaviour over longer intervals (depicted by spirals). Specifically, over time, these tendencies to quickly shift priorities according to social incentives can contribute to healthy exploration and risk-taking behaviours, which promote social and emotional learning and the development of skills and knowledge that underpin adult social competence. However, these same tendencies can also lead to negative spirals, such as when risk-taking and motivational learning processes respond to unhealthy incentives, such as drug and alcohol abuse or dangerous thrill-seeking. Another version of a negative spiral as a consequence of increased flexibility in adjusting goals and heightened sensitivity to social evaluation may be perceived failure in receiving admiration from peers, leading to disengagement from social goals, as seen in adolescent depression. The model proposes that changes in social–affective processing in combination with flexible prefrontal cortex (PFC) recruitment is generally adaptive and developmentally appropriate to the tasks and learning demands of adolescence, but in some situations — perhaps through interactions between individual risk factors and risk environments — can contribute to negative consequences. ACC, anterior cingulate cortex; DLPFC, dorsolateral PFC; mPFC, medial PFC; TPJ, temporoparietal junction.

This flexibility of cognitive control may also confer adaptive advantages for learning to navigate the often unpredictable social challenges of adolescence. The increased tendencies towards novelty-seeking and greater social–affective engagement might naturally nudge motivational tendencies towards the exploration of peer and romantic contexts. This may promote behavioural exploration in ways that create risks and vulnerabilities but also in ways that contribute to learning and developing new social–cognitive and social–affective skills. As described earlier, the fundamental task of adolescence is to achieve mature levels of social competence. The requisite skills require a great deal of practice, learning and refinement — particularly in the realms of self-control and affect regulation in socially charged situations. Natural tendencies to approach, explore and experiment with these often frightening, but sometimes thrilling, peer and romantic social situations — and to quickly engage

Box 5 | **An example of motivational flexibility in adolescent mice**

A study examined learning and decision-making in adolescent (or juvenile) (26–27-day-old) and adult (60–70-day-old) mice in a two-choice and four-choice odour-based foraging task[166]. The mice learned to discriminate different odours and learned which one was associated with a reward. Subsequently, the reward was paired with a different odour, and the reversal phase of the task assessed how fast the juvenile and adult mice learned this new association. The adolescent mice learned the four-choice discrimination and reversal faster than adult mice, with shorter choice latencies and more focused search strategies, suggestive of increased behavioural flexibility. The authors interpreted these findings as suggesting that adolescent mice are optimized to make flexible decisions in uncertain and unstable environments, which are likely to be encountered during adolescence.

frontal cortical systems in flexible ways — may promote key aspects of learning and social–affective development in adolescence.

The role of hormones on social–affective development. There has been growing interest in the cognitive, affective and social effects of puberty-related changes in the levels of several hormones, including oestradiol (which affects prefrontal functioning[145]), oxytocin (which influences social bonding and social motivation[146]), and adrenal androgens (dehydroepiandrosterone or dehydroepiandrosterone sulphate) and testosterone (which influence the motivation to attain and maintain social status[147–149]). Among these, we believe that the social effects of testosterone are particularly relevant to understanding some key changes in adolescence. Animal and human studies have shown that testosterone influences neural systems that regulate reward and social motivation. For example, in juvenile animals, testosterone has a crucial role in rough-and-tumble play, which serves as an important preparatory precursor to competition for dominance, territory maintenance and access to mates. Specifically, testosterone acts to direct attention and enhance approach to threatening social situations[150]. Data from human studies — including behavioural studies in which testosterone was administered to adults, experimental economic studies and functional neuroimaging studies — have provided compelling evidence for the role of testosterone as a social hormone[147–149]. Together, these findings indicate that testosterone promotes the search for and maintenance of social status, and that testosterone alters the appraisal of threats and rewards — particularly when these are relevant to social status[147,148]. A recent fMRI study in adults[151] showed that testosterone appears to cause a functional decoupling of amygdala and ventral PFC activity. The studies conducted in adults may be relevant to models of adolescent brain development because there is growing evidence for pubertal changes in ventral PFC, including the emergence of sex-differences at puberty[11].

To date, few studies have directly investigated how testosterone influences adolescent development. Preliminary findings from fMRI studies suggest that testosterone levels correlate with maturational changes in reward processing in adolescent boys and girls[152,153]. Structural MRI studies have shown associations between circulating testosterone levels and cortical thickness in the left inferior parietal lobule, middle temporal gyrus, calcarine sulcus and right lingual gyrus, which are all regions known to be high in androgen receptors. Of note, however, the fMRI findings show similar testosterone effects on male and female reward processing, whereas the structural findings showed sex differences, with testosterone being associated with grey matter thinning in girls but with grey matter thickening in boys[154].

There is a need for a better understanding of the effects of testosterone (and other hormones) on behaviour and brain function during human adolescent development. The evidence for the role of testosterone in social motivation (in animal studies and studies in adult humans) raises compelling questions about the role of testosterone in social–affective changes during adolescence. For example, if the pubertal surge in testosterone levels amplifies the motivational salience of social status (in both boys and girls), adolescents may show a general increase in the motivation to be admired. The specific types of behaviour (and reward learning) that result from this increased motivation could vary widely across cultural contexts. Thus, in a culture that admires bold, assertive behaviour in boys but not in girls, different adolescent experiences in boys versus girls may sculpt motivational learning in fundamentally different ways through patterns of adolescent experience. Similarly, in a Tibetan Buddhist monastery, where adolescent boys may be competing for social status by demonstrating the greatest kindness and compassion, the testosterone-amplified desire to be admired might promote a very different pattern of motivational learning in boys than in other societies. These examples highlight the importance of interactions between biology and social context in the refinement of neural circuitry in adolescence.

Conclusions and future directions

As highlighted in this Review, some of the most compelling questions about the adolescent window of maturation focus on the affective dimension of motivations and goals. This includes mechanistic questions about hormone-specific effects in early adolescence that contribute to the intensification of feelings related to social valuation. Progress in understanding these mechanistic questions may provide insights into the unique opportunities for motivational learning in adolescence. For example, how do social and affective learning in adolescence contribute to the development of individual differences in motivational priorities, such as enduring heartfelt goals? It seems clear that these learning processes involve implicit and affective aspects of developing one's values and attitudes as well as the explicit cognitive processes of setting priorities. For example, individual differences in the tendencies to be kind, honest and loyal in a romantic relationship may have as much to do with one's feelings about these values as with consciously weighed decisions about the consequences of such behaviours. Another example concerns acquired intrinsic motivations in adolescence. Progress in identifying the neurodevelopmental underpinnings of these acquired motivations are relevant to understanding the development of healthy versions of inspired passions as well as vulnerabilities for developing unhealthy versions

of acquired motivations, such as drug and alcohol use and reckless versions of thrill-seeking.

It seems likely that during several phases of development across the lifespan, neural systems in the PFC may have some 'experience-expectant' qualities — that is, they may have windows of development during which the brain 'expects' or is biologically prepared for learning. These qualities enable adaptive adjustments that are relevant to the challenges and opportunities that tended to occur at that phase of development during our evolutionary history. Accordingly, the social challenges and changes facing adolescents (throughout human history[155]) may have favoured a slightly different cognitive style (more flexible, exploratory and sensitive to social–affective influences) compared with adults. This notion argues against the idea that the adult brain is the optimal or 'normal' functional system and that differences during adolescent development represent 'deficits'.

As we have described in this Review, there is a compelling need for studies that advance, refine and test key features of this heuristic model at the level of the underlying neural changes, in large part because these questions have such relevance to early intervention and prevention for a wide range of adolescent-onset health problems, as well as broad implications for health, education, juvenile justice and social policies aimed at youths.

1. Dahl, R. E. & Gunnar, M. R. Heightened stress responsiveness and emotional reactivity during pubertal maturation: implications for psychopathology. *Dev. Psychopathol.* **21**, 1–6 (2009).
2. Steinberg, L. A. Social neuroscience perspective on adolescent risk-taking. *Dev. Rev.* **28**, 78–106 (2008).
3. Blakemore, S. J., Burnett, S. & Dahl, R. E. The role of puberty in the developing adolescent brain. *Hum. Brain Mapp.* **31**, 926–933 (2010).
4. Arnett, J. J. *Adolescence and Emerging Adulthood: A Cultural Approach* (Prentice Hall, 2004).
5. Dahl, R. E. & Vanderschuren, L. J. The feeling of motivation in the developing brain. *Dev. Cogn. Neurosci.* **1**, 361–363 (2011).
6. Gladwin, T. E., Figner, B., Crone, E. A. & Wiers, R. W. Addiction, adolescence, and the integration of control and motivation. *Dev. Cogn. Neurosci.* **1**, 364–376 (2011).
7. Somerville, L. H., Jones, R. M. & Casey, B. J. A time of change: behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain Cogn.* **72**, 124–133 (2010).
An influential paper that describes the dual processing model, suggesting that adolescence is characterized by faster maturation of subcortical brain regions relative to frontal cortical regions.
8. Ernst, M. & Fudge, J. L. A developmental neurobiological model of motivated behavior: anatomy, connectivity and ontogeny of the triadic nodes. *Neurosci. Biobehav. Rev.* **33**, 367–382 (2009).
9. Steinberg, L. *et al.* Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: evidence for a dual systems model. *Dev. Psychol.* **44**, 1764–1778 (2008).
10. Nelson, E. E., Leibenluft, E., McClure, E. B. & Pine, D. S. The social re-orientation of adolescence: a neuroscience perspective on the process and its relation to psychopathology. *Psychol. Med.* **35**, 163–174 (2005).
11. Nelson, E. E. & Guyer, A. E. The development of the ventral prefrontal cortex and social flexibility. *Dev. Cogn. Neurosci.* **1**, 233–245 (2011).
A very interesting paper that focuses on the role of ventral prefrontal circuitry and social flexibility in adolescent development.
12. Pfeifer, J. H. & Allen, N. B. Arrested development? Reconsidering dual-systems models of brain function in adolescence and disorders. *Trends Cogn. Sci.* **16**, 322–329 (2012).
13. Asato, M. R., Sweeney, J. A. & Luna, B. Cognitive processes in the development of TOL performance. *Neuropsychologia* **44**, 2259–2269 (2006).
14. Huizinga, M., Dolan, C. V. & van der Molen, M. W. Age-related change in executive function: developmental trends and a latent variable analysis. *Neuropsychologia* **44**, 2017–2036 (2006).
15. Case, R. *The Mind's Staircase: Exploring the Conceptual Underpinnings of Children's Thought and Knowledge* (Erlbaum, 1992).
16. Zelazo, P. D., Craik, F. I. & Booth, L. Executive function across the life span. *Acta Psychol. (Amst.)* **115**, 167–183 (2004).
17. Zelazo, P. D. The development of conscious control in childhood. *Trends Cogn. Sci.* **8**, 12–17 (2004).
18. Miyake, A. *et al.* The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cogn. Psychol.* **41**, 49–100 (2000).
19. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
20. Kwon, H., Reiss, A. L. & Menon, V. Neural basis of protracted developmental changes in visuo-spatial working memory. *Proc. Natl Acad. Sci. USA* **99**, 13336–13341 (2002).
21. Klingberg, T., Forssberg, H. & Westerberg, H. Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *J. Cogn. Neurosci.* **14**, 1–10 (2002).
22. Schweinsburg, A. D., Nagel, B. J. & Tapert, S. F. fMRI reveals alteration of spatial working memory networks across adolescence. *J. Int. Neuropsychol. Soc.* **11**, 631–644 (2005).
23. Scherf, K. S., Sweeney, J. A. & Luna, B. Brain basis of developmental change in visuospatial working memory. *J. Cogn. Neurosci.* **18**, 1045–1058 (2006).
24. Crone, E. A., Wendelken, C., Donohue, S., van Leijenhorst, L. & Bunge, S. A. Neurocognitive development of the ability to manipulate information in working memory. *Proc. Natl Acad. Sci. USA* **103**, 9315–9320 (2006).
25. Ciesielski, K. T., Lesnik, P. G., Savoy, R. L., Grant, E. P. & Ahlfors, S. P. Developmental neural networks in children performing a Categorical *N*-Back Task. *Neuroimage* **33**, 980–990 (2006).
26. Olesen, P. J., Macoveanu, J., Tegner, J. & Klingberg, T. Brain activity related to working memory and distraction in children and adults. *Cereb. Cortex* **17**, 1047–1054 (2007).
27. Thomason, M. E. *et al.* Development of spatial and verbal working memory capacity in the human brain. *J. Cogn. Neurosci.* **21**, 316–332 (2009).
28. O'Hare, E. D., Lu, L. H., Houston, S. M., Bookheimer, S. Y. & Sowell, E. R. Neurodevelopmental changes in verbal working memory load-dependency: an fMRI investigation. *Neuroimage* **42**, 1678–1685 (2008).
29. Jolles, D. D., Kleibeuker, S. W., Rombouts, S. A. & Crone, E. A. Developmental differences in prefrontal activation during working memory maintenance and manipulation for different memory loads. *Dev. Sci.* **14**, 713–724 (2011).
30. Wendelken, C., Baym, C. L., Gazzaley, A. & Bunge, S. A. Neural indices of improved attentional modulation over middle childhood. *Dev. Cogn. Neurosci.* **1**, 175–186 (2011).
31. Adleman, N. E. *et al.* A developmental fMRI study of the Stroop color-word task. *Neuroimage* **16**, 61–75 (2002).
32. Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J. & Gabrieli, J. D. Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. *Neuron* **33**, 301–311 (2002).
33. Casey, B. J., Thomas, K. M., Davidson, M. C., Kunz, K. & Franzen, P. L. Dissociating striatal and hippocampal function developmentally with a stimulus–response compatibility task. *J. Neurosci.* **22**, 8647–8652 (2002).
34. Marsh, R. *et al.* A developmental fMRI study of self-regulatory control. *Hum. Brain Mapp.* **27**, 848–863 (2006).
35. Rubia, K. *et al.* Progressive increase of frontostriatal brain activation from childhood to adulthood during event-related tasks of cognitive control. *Hum. Brain Mapp.* **27**, 973–993 (2006).
36. Rubia, K., Smith, A. B., Taylor, E. & Brammer, M. Linear age-correlated functional development of right inferior fronto-striato-cerebellar networks during response inhibition and anterior cingulate during error-related processes. *Hum. Brain Mapp.* **28**, 1163–1177 (2007).
37. Casey, B. J. *et al.* Early development of subcortical regions involved in non-cued attention switching. *Dev. Sci.* **7**, 534–542 (2004).
38. Christakou, A. *et al.* Sex-dependent age modulation of frontostriatal and temporoparietal activation during cognitive control. *Neuroimage* **48**, 223–236 (2009).
39. Crone, E. A., Donohue, S. E., Honomichl, R., Wendelken, C. & Bunge, S. A. Brain regions mediating flexible rule use during development. *J. Neurosci.* **26**, 11239–11247 (2006).
40. Bunge, S. A. & Wright, S. B. Neurodevelopmental changes in working memory and cognitive control. *Curr. Opin. Neurobiol.* **17**, 243–250 (2007).
41. Geier, C. F., Garver, K., Terwilliger, R. & Luna, B. Development of working memory maintenance. *J. Neurophysiol.* **101**, 84–99 (2009).
42. Brahmbhatt, S. B., McAuley, T. & Barch, D. M. Functional developmental similarities and differences in the neural correlates of verbal and nonverbal working memory tasks. *Neuropsychologia* **46**, 1020–1031 (2008).
43. Libertus, M. E., Brannon, E. M. & Pelphrey, K. A. Developmental changes in category-specific brain responses to numbers and letters in a working memory task. *Neuroimage* **44**, 1404–1414 (2009).
44. Tamm, L., Menon, V. & Reiss, A. L. Maturation of brain function associated with response inhibition. *J. Am. Acad. Child Adolesc. Psychiatry* **41**, 1231–1238 (2002).
45. Durston, S. *et al.* A shift from diffuse to focal cortical activity with development. *Dev. Sci.* **9**, 1–8 (2006).
46. Booth, J. R. *et al.* Neural development of selective attention and response inhibition. *Neuroimage* **20**, 737–751 (2003).
47. Velanova, K., Wheeler, M. E. & Luna, B. The maturation of task set-related activation supports late developmental improvements in inhibitory control. *J. Neurosci.* **29**, 12558–12567 (2009).
48. Cohen, J. R. *et al.* Decoding developmental differences and individual variability in response inhibition through predictive analyses across individuals. *Front. Hum. Neurosci.* **4**, 47 (2010).
49. Konrad, K. *et al.* Development of attentional networks: an fMRI study with children and adults. *Neuroimage* **28**, 429–439 (2005).
50. Morton, J. B., Bosma, R. & Ansari, D. Age-related changes in brain activation associated with dimensional shifts of attention: an fMRI study. *Neuroimage* **46**, 249–256 (2009).
51. Luna, B. *et al.* Maturation of widely distributed brain function subserves cognitive development. *Neuroimage* **13**, 786–793 (2001).
52. Brahmbhatt, S. B., White, D. A. & Barch, D. M. Developmental differences in sustained and transient activity underlying working memory. *Brain Res.* **1354**, 140–151 (2010).
53. Johnson, M. H., Grossmann, T. & Cohen Kadosh, K. Mapping functional brain development: building a social brain through interactive specialization. *Dev. Psychol.* **45**, 151–159 (2009).
54. Luna, B., Padmanabhan, A. & O'Hearn, K. What has fMRI told us about the development of cognitive control through adolescence? *Brain Cogn.* **72**, 101–113 (2010).

55. Crone, E. A., Zanolie, K., Van Leijenhorst, L., Westenberg, P. M. & Rombouts, S. A. Neural mechanisms supporting flexible performance adjustment during development. *Cogn. Affect. Behav. Neurosci.* **8**, 165–177 (2008).
56. Cohen, J. R. *et al.* A unique adolescent response to reward prediction errors. *Nature Neurosci.* **13**, 669–671 (2010).
One of the first studies investigating the development of the prediction error in adolescents. It describes how this may underpin some changes in risk-taking in adolescence.
57. van den Bos, W., Guroglu, B., van den Bulk, B. G., Rombouts, S. A. & Crone, E. A. Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. *Front. Hum. Neurosci.* **3**, 52 (2009).
58. van Duijvenvoorde, A. C., Zanolie, K., Rombouts, S. A., Raijmakers, M. E. & Crone, E. A. Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *J. Neurosci.* **28**, 9495–9503 (2008).
59. Velanova, K., Wheeler, M. E. & Luna, B. Maturational changes in anterior cingulate and frontoparietal recruitment support the development of error processing and inhibitory control. *Cereb. Cortex* **18**, 2505–2522 (2008).
60. Dumontheil, I., Houlton, R., Christoff, K. & Blakemore, S. J. Development of relational reasoning during adolescence. *Dev. Sci.* **13**, F15–F24 (2010).
61. Dumontheil, I., Hassan, B., Gilbert, S. J. & Blakemore, S. J. Development of the selection and manipulation of self-generated thoughts in adolescence. *J. Neurosci.* **30**, 7664–7671 (2010).
62. Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E. & Bunge, S. A. Neural correlates of fluid reasoning in children and adults. *Front. Hum. Neurosci.* **1**, 8 (2007).
63. Crone, E. A. *et al.* Neurocognitive development of relational reasoning. *Dev. Sci.* **12**, 55–66 (2009).
64. Fuligni, A. J. in *Research on Minority Adolescents: Conceptual, Theoretical, and Methodological Issues* (eds McLoyd, V. & Steinberg, L.) 127–143 (Erlbaum, 1998).
65. Finn, A. S., Sheridan, M. A., Kam, C. L., Hinshaw, S. & D'Esposito, M. Longitudinal evidence for functional specialization of the neural circuit supporting working memory in the human brain. *J. Neurosci.* **30**, 11062–11067 (2010).
66. Koolschijn, P. C., Schel, M. A., de Rooij, M., Rombouts, S. A. & Crone, E. A. A three-year longitudinal functional magnetic resonance imaging study of performance monitoring and test-retest reliability from childhood to early adulthood. *J. Neurosci.* **31**, 4204–4212 (2011).
67. Jolles, D. D., van Buchem, M. A., Rombouts, S. A. & Crone, E. A. Practice effects in the developing brain: a pilot study. *Dev. Cogn. Neurosci.* **2** (Suppl. 1), 180–191 (2012).
68. Jolles, D. D., van Buchem, M. A., Crone, E. A. & Rombouts, S. A. A comprehensive study of whole-brain functional connectivity in children and young adults. *Cereb. Cortex* **21**, 385–391 (2011).
69. Fair, D. A. *et al.* Development of distinct control networks through segregation and integration. *Proc. Natl Acad. Sci. USA* **104**, 13507–13512 (2007).
A thoughtful analysis of the development of resting state networks from childhood through adolescence.
70. Jolles, D. D., van Buchem, M. A., Crone, E. A. & Rombouts, S. A. Functional brain connectivity at rest changes after working memory training. *Hum. Brain Mapp.* 11 Nov 2011 (doi:10.1002/hbm.21444).
71. Haber, S. N. & Knutson, B. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* **35**, 4–26 (2010).
72. Ernst, M. *et al.* Amygdala and nucleus accumbens in responses to receipt and omission of gains in adults and adolescents. *Neuroimage* **25**, 1279–1291 (2005).
73. Van Leijenhorst, L. *et al.* Adolescent risky decision-making: neurocognitive development of reward and control regions. *Neuroimage* **51**, 345–355 (2010).
74. Van Leijenhorst, L. *et al.* What motivates the adolescent? Brain regions mediating reward sensitivity across adolescence. *Cereb. Cortex* **20**, 61–69 (2010).
75. Galvan, A. *et al.* Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. *J. Neurosci.* **26**, 6885–6892 (2006).
76. Geier, C. F., Terwilliger, R., Teslovich, T., Velanova, K. & Luna, B. Immaturities in reward processing and its influence on inhibitory control in adolescence. *Cereb. Cortex* **20**, 1613–1629 (2010).
One of the first studies to demonstrate that incentives appear to have a particularly pronounced effect on cognitive control in adolescents.
77. Chein, J., Albert, D., O'Brien, L., Uckert, K. & Steinberg, L. Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Dev. Sci.* **14**, F1–F10 (2011).
78. Smith, A. B., Halari, R., Giampetro, V., Brammer, M. & Rubia, K. Developmental effects of reward on sustained attention networks. *Neuroimage* **56**, 1693–1704 (2011).
79. Christakou, A., Brammer, M. & Rubia, K. Maturation of limbic corticostriatal activation and connectivity associated with developmental changes in temporal discounting. *Neuroimage* **54**, 1344–1354 (2011).
80. Padmanabhan, A., Geier, C. F., Ordaz, S. J., Teslovich, T. & Luna, B. Developmental changes in brain function underlying the influence of reward processing on inhibitory control. *Dev. Cogn. Neurosci.* **1**, 517–529 (2011).
81. Bjork, J. M. *et al.* Incentive-elicited brain activation in adolescents: similarities and differences from young adults. *J. Neurosci.* **24**, 1793–1802 (2004).
82. Bjork, J. M., Smith, A. R., Chen, G. & Hommer, D. W. Adolescents, adults and rewards: comparing motivational neurocircuitry recruitment using fMRI. *PLoS ONE* **5**, e11440 (2010).
83. Forbes, E. E. & Dahl, R. E. Pubertal development and behavior: hormonal activation of social and motivational tendencies. *Brain Cogn.* **72**, 66–72 (2010).
84. Krain, A. L. *et al.* An fMRI examination of developmental differences in the neural correlates of uncertainty and decision-making. *J. Child Psychol. Psychiatry* **47**, 1023–1030 (2006).
85. Eshel, N., Nelson, E. E., Blair, R. J., Pine, D. S. & Ernst, M. Neural substrates of choice selection in adults and adolescents: development of the ventrolateral prefrontal and anterior cingulate cortices. *Neuropsychologia* **45**, 1270–1279 (2007).
86. Bjork, J. M., Smith, A. R., Danube, C. L. & Hommer, D. W. Developmental differences in posterior mesofrontal cortex recruitment by risky rewards. *J. Neurosci.* **27**, 4839–4849 (2007).
87. Van den Bos, W., Cohen, M. X., Kahnt, T. & Crone, E. A. Striatum–medial prefrontal cortex connectivity predicts developmental differences in reinforcement learning. *Cereb. Cortex* **22**,1247–1255 (2012).
88. Sugam, J. A., Day, J. J., Wightman, R. M. & Carelli, R. M. Phasic nucleus accumbens dopamine encodes risk-based decision-making behavior. *Biol. Psychiatry* **71**, 199–205 (2012).
89. Killgore, W. D., Oki, M. & Yurgelun-Todd, D. A. Sex-specific developmental changes in amygdala responses to affective faces. *Neuroreport* **12**, 427–433 (2001).
90. Monk, C. S. *et al.* Adolescent immaturity in attention-related brain engagement to emotional facial expressions. *Neuroimage* **20**, 420–428 (2003).
91. Williams, L. M. *et al.* The mellow years?: neural basis of improving emotional stability over age. *J. Neurosci.* **26**, 6422–6430 (2006).
92. Guyer, A. E. *et al.* A developmental examination of amygdala response to facial expressions. *J. Cogn. Neurosci.* **20**, 1565–1582 (2008).
93. Hare, T. A. *et al.* Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. *Biol. Psychiatry* **63**, 927–934 (2008).
94. Pfeifer, J. H. *et al.* Entering adolescence: resistance to peer influence, risky behavior, and neural changes in emotion reactivity. *Neuron* **69**, 1029–1036 (2011).
One of the first studies showing longitudinally that increases in striatum response may be adaptive.
95. Pine, D. S. *et al.* Cortical brain regions engaged by masked emotional faces in adolescents and adults: an fMRI study. *Emotion* **1**, 137–147 (2001).
96. Thomas, K. M. *et al.* Amygdala response to facial expressions in children and adults. *Biol. Psychiatry* **49**, 309–316 (2001).
97. Nelson, E. E. *et al.* Developmental differences in neuronal engagement during implicit encoding of emotional faces: an event-related fMRI study. *J. Child Psychol. Psychiatry* **44**, 1015–1024 (2003).
98. Yurgelun-Todd, D. A. & Killgore, W. D. Fear-related activity in the prefrontal cortex increases with age during adolescence: a preliminary fMRI study. *Neurosci. Lett.* **406**, 194–199 (2006).
99. Deeley, Q. *et al.* Changes in male brain responses to emotional faces from adolescence to middle age. *Neuroimage* **40**, 389–397 (2008).
100. Somerville, L. H., Hare, T. & Casey, B. J. Frontostriatal maturation predicts cognitive control failure to appetitive cues in adolescents. *J. Cogn. Neurosci.* **23**, 2123–2134 (2011).
101. Casey, B., Jones, R. M. & Somerville, L. H. Braking and accelerating of the adolescent brain. *J. Res. Adolesc.* **21**, 21–33 (2011).
102. Schaffer, H. R. *Social Development* (Blackwell,1996).
103. Steinberg, L. & Morris, A. S. Adolescent development. *Annu. Rev. Psychol.* **52**, 83–110 (2001).
104. Burnett, S., Sebastian, C., Cohen Kadosh, K. & Blakemore, S. J. The social brain in adolescence: evidence from functional magnetic resonance imaging and behavioural studies. *Neurosci. Biobehav. Rev.* **35**, 1654–1664 (2011).
105. Rilling, J. K. & Sanfey, A. G. The neuroscience of social decision-making. *Annu. Rev. Psychol.* **62**, 23–48 (2011).
106. Blakemore, S. J. The social brain in adolescence. *Nature Rev. Neurosci.* **9**, 267–277 (2008).
107. Wang, A. T., Lee, S. S., Sigman, M. & Dapretto, M. Developmental changes in the neural basis of interpreting communicative intent. *Soc. Cogn. Affect. Neurosci.* **1**, 107–121 (2006).
108. Blakemore, S. J., den Ouden, H., Choudhury, S. & Frith, C. Adolescent development of the neural circuitry for thinking about intentions. *Soc. Cogn. Affect. Neurosci.* **2**, 130–139 (2007).
109. Moriguchi, Y., Ohnishi, T., Mori, T., Matsuda, H. & Komaki, G. Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence. *Psychiatry Clin. Neurosci.* **61**, 355–363 (2007).
110. Kobayashi, C., Glover, G. H. & Temple, E. Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. *Neuropsychologia* **45**, 1522–1532 (2007).
111. Burnett, S., Bird, G., Moll, J., Frith, C. & Blakemore, S. J. Development during adolescence of the neural processing of social emotion. *J. Cogn. Neurosci.* **21**, 1736–1750 (2009).
112. Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J. & Pelphrey, K. A. Brain regions for perceiving and reasoning about other people in school-aged children. *Child Dev.* **80**, 1197–1209 (2009).
113. Gunther Moor, B. *et al.* Neurodevelopmental changes of reading the mind in the eyes. *Soc. Cogn. Affect. Neurosci.* **7**, 44–52 (2012).
114. Sebastian, C. L. *et al.* Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Soc. Cogn. Affect. Neurosci.* **7**, 53–63 (2012).
115. Pfeifer, J. H., Lieberman, M. D. & Dapretto, M. "I know you are but what am I?!": neural bases of self- and social knowledge retrieval in children and adults. *J. Cogn. Neurosci.* **19**, 1323–1337 (2007).
116. Pfeifer, J. H. *et al.* Neural correlates of direct and reflected self-appraisals in adolescents and adults: when social perspective-taking informs self-perception. *Child Dev.* **80**, 1016–1038 (2009).
117. Eisenberg, N. & Fabes, F. in *Handbook of Child Psychology: Social, Emotional, and Personality Development* (eds Damon, W., Eisenberg, N. & Lerner, R.) 646–718 (Wiley, 2006).
118. Newcomb, A. F., Bukowski, W. M. & Pattee, L. Children's peer relations: a meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychol. Bull.* **113**, 99–128 (1993).
119. Güth, W., Schmittberger, R. & Schwarze, B. An experimental analysis of ultimatum bargaining. *J. Econom. Behav. Organiz.* **3**, 367 (1982).
120. Berg, J., Dickhaut, J. & McCabe, K. Trust, reciprocity, and social history. *Games Econom. Behav.* **10**, 122–142 (1995).
121. Guroglu, B., van den Bos, W. & Crone, E. A. Fairness considerations: increasing understanding of intentionality during adolescence. *J. Exp. Child Psychol.* **104**, 398–409 (2009).
122. van den Bos, W., Westenberg, M., van Dijk, E. & Crone, E. A. Development of trust and reciprocity in adolescence. *Cogn. Dev.* **25**, 90–102 (2010).
123. Steinbeis, N., Bernhardt, B. C. & Singer, T. Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. *Neuron* **73**, 1040–1051 (2012).

124. Van Overwalle, F. Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* **30**, 829–858 (2009).
125. van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. & Crone, E. A. What motivates repayment? Neural correlates of reciprocity in the Trust Game. *Soc. Cogn. Affect. Neurosci.* **4**, 294–304 (2009).
126. Guroglu, B., van den Bos, W., Rombouts, S. A. & Crone, E. A. Unfair? It depends: neural correlates of fairness in social context. *Soc. Cogn. Affect. Neurosci.* **5**, 414–423 (2010).
127. Guroglu, B., van den Bos, W., van Dijk, E., Rombouts, S. A. & Crone, E. A. Dissociable brain networks involved in development of fairness considerations: understanding intentionality behind unfairness. *Neuroimage* **57**, 634–641 (2011).
128. van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. & Crone, E. A. Changing brains, changing perspectives: the neurocognitive development of reciprocity. *Psychol. Sci.* **22**, 60–70 (2011).
A social interaction study that shows, using a neuroeconomics approach, a transition from self-referential processing to other-referential processing in adolescence.
129. Cillessen, A. H. & Rose, A. J. Understanding popularity in the peer system. *Curr. Direct. Psychol. Sci.* **14**, 102–105 (2005).
130. Decety, J. & Michalska, K. J. Neurodevelopmental changes in the circuits underlying empathy and sympathy from childhood to adulthood. *Dev. Sci.* **13**, 886–899 (2010).
131. Gunther Moor, B., van Leijenhorst, L., Rombouts, S. A., Crone, E. A. & Van der Molen, M. W. Do you like me? Neural correlates of social evaluation and developmental trajectories. *Soc. Neurosci.* **5**, 461–482 (2010).
132. Guyer, A. E., McClure-Tone, E. B., Shiffrin, N. D., Pine, D. S. & Nelson, E. E. Probing the neural correlates of anticipated peer evaluation in adolescence. *Child Dev.* **80**, 1000–1015 (2009).
133. Sebastian, C. L. *et al.* Effects of age and MAOA genotype on the neural processing of social rejection. *Genes Brain Behav.* **9**, 628–637 (2010).
134. Sebastian, C. L. *et al.* Developmental influences on the neural bases of responses to social rejection: implications of social neuroscience for education. *Neuroimage* **57**, 686–694 (2011).
135. Gunther Moor, B. *et al.* Social exclusion and punishment of excluders: neural correlates and developmental trajectories. *Neuroimage* **59**, 708–717 (2012).
On the basis of a wide age range of participants, this is one of the first studies showing that social rejection in adolescence leads to subsequent punishment of excluders.
136. Masten, C. L. *et al.* Neural correlates of social exclusion during adolescence: understanding the distress of peer rejection. *Soc. Cogn. Affect. Neurosci.* **4**, 143–157 (2009).
137. Masten, C. L., Telzer, E. H., Fuligni, A. J., Lieberman, M. D. & Eisenberger, N. I. Time spent with friends in adolescence relates to less neural sensitivity to later peer rejection. *Soc. Cogn. Affect. Neurosci.* **7**, 106–114 (2012).
138. Masten, C. L. *et al.* Subgenual anterior cingulate responses to peer rejection: a marker of adolescents' risk for depression. *Dev. Psychopathol.* **23**, 283–292 (2011).
139. Angold, A., Costello, E. J. & Worthman, C. M. Puberty and depression: the roles of age, pubertal status, and pubertal timing. *Psychol. Med.* **28**, 51–61 (1998).
140. Dahl, R. E. Adolescent brain development: a period of vulnerabilities and opportunities. *Ann. NY Acad. Sci.* **1021**, 1–22 (2004).
141. Galvan, A. Adolescent development of the reward system. *Front. Hum. Neurosci.* **4**, 6 (2010).
142. Steinberg, L. *The Science of Adolescent Risk-Taking.* (Washington, 2011).
143. Figner, B., Mackinlay, R. J., Wilkening, F. & Weber, E. U. Affective and deliberative processes in risky choice: age differences in risk taking in the Columbia Card Task. *J. Exp. Psychol. Learn. Mem. Cogn.* **35**, 709–730 (2009).
144. Kleibeuker, S. W., De Dreu, C. K. W. & Crone, E. A. The development of creative cognition across adolescence: distinct trajectories for insight and divergent thinking. *Dev. Sci.* (in the press).
145. Jacobs, E. & D'Esposito, M. Estrogen shapes dopamine-dependent cognitive processes: implications for women's health. *J. Neurosci.* **31**, 5286–5293 (2011).
146. Gordon, I., Martin, C., Feldman, R. & Leckman, J. F. Oxytocin and social motivation. *Dev. Cogn. Neurosci.* **1**, 471–493 (2011).
147. Bos, P. A., Panksepp, J., Bluthe, R. M. & van Honk, J. Acute effects of steroid hormones and neuropeptides on human social–emotional behavior: a review of single administration studies. *Front. Neuroendocrinol.* **33**, 17–35 (2012).
148. Eisenegger, C., Haushofer, J. & Fehr, E. The role of testosterone in social interaction. *Trends Cogn. Sci.* **15**, 263–271 (2010).
A compelling review on the influence of testosterone as a social hormone that influences social information processing and motivation.
149. Carney, D. & Mason, M. F. Decision making and testosterone: when the ends justify the means. *J. Exp. Social Psychol.* **46**, 668–671 (2010).
150. Cooke, B. M. & Shukla, D. Double helix: Reciprocity between juvenile play and brain development. *Dev. Cogn. Neurosci.* **1**, 459–470 (2011).
151. Van Wingen, G., Mattern, C., Verkes, R. J., Buitelaar, J. & Fernandez, G. Testosterone reduces amygdala–orbitofrontal cortex coupling. *Psychoneuroendocrinology* **35**, 105–113 (2010).
152. Forbes, E. E. *et al.* Healthy adolescents' neural response to reward: associations with puberty, positive affect, and depressive symptoms. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 162–172 (2010).
153. Op de Macks, Z. *et al.* Testosterone levels correspond with increased ventral striatum activation in response to monetary rewards in adolescents. *Dev. Cogn. Neurosci.* **1**, 506–516 (2011).
154. Bramen, J. E. *et al.* Sex matters during adolescence: testosterone-related cortical thickness maturation differs between boys and girls. *PLoS ONE* **7**, e33850 (2012).
155. Schlegel, A. & Barry, H. *Adolescence: An Anthropological Inquiry* (Free Press, 1991).
156. Ross, J., Roeltgen, D. & Zinn, A. Cognition and the sex chromosomes: studies in Turner syndrome. *Horm. Res.* **65**, 47–56 (2006).
157. Sowell, E. R. *et al.* Longitudinal mapping of cortical thickness and brain growth in normal children. *J. Neurosci.* **24**, 8223–8231 (2004).
158. Gogtay, N. *et al.* Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl Acad. Sci. USA* **101**, 8174–8179 (2004).
159. Giedd, J. N. Structural magnetic resonance imaging of the adolescent brain. *Ann. NY Acad. Sci.* **1021**, 77–85 (2004).
160. Lenroot, R. K. & Giedd, J. N. Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neurosci. Biobehav. Rev.* **30**, 718–729 (2006).
161. Ostby, Y. *et al.* Heterogeneity in subcortical brain development: a structural magnetic resonance imaging study of brain maturation from 8 to 30 years. *J. Neurosci.* **29**, 11772–11782 (2009).
162. Spear, L. P. Heightened stress responsivity and emotional reactivity during pubertal maturation: implications for psychopathology. *Dev. Psychopathol.* **21**, 87–97 (2009).
163. Sisk, C. L. & Zehr, J. L. Pubertal hormones organize the adolescent brain and behavior. *Front. Neuroendocrinol.* **26**, 163–174 (2005).
164. Galvan, A. Neural plasticity of development and learning. *Hum. Brain Mapp.* **31**, 879–890 (2010).
165. Wahlstrom, D., White, T. & Luciana, M. Neurobehavioral evidence for changes in dopamine system activity during adolescence. *Neurosci. Biobehav. Rev.* **34**, 631–648 (2010).
166. Johnson, C. & Wilbrecht, L. Juvenile mice show greater flexibility in multiple choice reversal learning than adults. *Dev. Cogn. Neurosci.* **1**, 540–551 (2011).
167. Olesen, P. J., Nagy, Z., Westerberg, H. & Klingberg, T. Combined analysis of DTI and fMRI data reveals a joint maturation of white and grey matter in a fronto-parietal network. *Brain Res. Cogn. Brain Res.* **18**, 48–57 (2003).
168. Durston, S. *et al.* A neural basis for the development of inhibitory control. *Dev. Sci.* **5**, F9–F16 (2002).
169. May, J. C. *et al.* Event-related functional magnetic resonance imaging of reward-related brain circuitry in children and adolescents. *Biol. Psychiatry* **55**, 359–366 (2004).
170. Guyer, A. E. *et al.* Amygdala and ventrolateral prefrontal cortex function during anticipated peer evaluation in pediatric social anxiety. *Arch. Gen. Psychiatry* **65**, 1303–1312 (2008).
171. Decety, J., Michalska, K. J. & Kinzler, K. D. The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study. *Cereb. Cortex* **22**, 209–220 (2012).
172. Harenski, C. L., Harenski, K. A., Shane, M. S. & Kiehl, K. A. Neural development of mentalizing in moral judgment from adolescence to adulthood. *Dev. Cogn. Neurosci.* **2**, 162–173 (2012).
173. Church, J. A., Petersen, S. E. & Schlaggar, B. L. The "Task B problem" and other considerations in developmental functional neuroimaging. *Hum. Brain Mapp.* **31**, 852–862 (2010).

Acknowledgements
E.A.C. is supported by grants from the European Research Counsil (ERC), the Netherlands Science Foundation (NWO) and the Young Academy of the Royal Netherlands Academy of Arts and Sciences. R.E.D is supported by grants from the National Institute of Mental Health, National Institute of Drug Abuse, National Institute of Child Health and Human Development and National Institute on Alcohol Abuse and Alcoholism.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Eveline A. Crone's homepage: www.brainanddevelopmentlab.nl

SUPPLEMENTARY INFORMATION
See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Molecular Psychiatry (2014) 19, 791–800

www.nature.com/mp

ORIGINAL ARTICLE

Serum BDNF concentrations as peripheral manifestations of depression: evidence from a systematic review and meta-analyses on 179 associations ($N = 9484$)

ML Molendijk[1,2], P Spinhoven[1,2,3], M Polak[1], BAA Bus[4], BWJH Penninx[3,5,6] and BM Elzinga[1,2]

Meta-analyses, published in 2008–2010, have confirmed abnormally low serum brain-derived neurotrophic factor (BDNF) concentrations in depressed patients and normalization of this by antidepressant treatment. These findings are believed to reflect peripheral manifestations of the *neurotrophin hypothesis*, which states that depression is secondary to an altered expression of BDNF in the brain. Since the publication of these meta-analyses, the field has seen a huge increase in studies on these topics. This motivated us to update the evidence on the aforementioned associations and, in addition, to compile the data on serum BDNF concentrations in relation to the symptom severity of depression. Using a manifold of data as compared with earlier meta-analyses, we find low serum BDNF concentrations in 2384 antidepressant-free depressed patients relative to 2982 healthy controls and to 1249 antidepressant-treated depressed patients (Cohen's $d = -0.71$ and -0.56, P-values < 0.0000001). When publication bias is accounted for, these effect-sizes become substantially smaller ($d = -0.47$ and -0.34, respectively, P-values < 0.0001). We detect between-study heterogeneity in outcomes for which only year of publication and sample size are significant moderators, with more recent papers and larger samples sizes in general being associated with smaller between-group differences. Finally, the aggregated data negate consistent associations between serum BDNF concentrations and the symptom severity of depression. Our findings corroborate the claim that altered serum BDNF concentrations are peripheral manifestations of depression. However, here we highlight that the evidence for this claim is slimmer as was initially thought and amidst a lot of noise.

Molecular Psychiatry (2014) **19,** 791–800; doi:10.1038/mp.2013.105; published online 20 August 2013

Keywords: antidepressants; BDNF; biomarker; depression; meta-analysis

INTRODUCTION

The *neurotrophin hypothesis*, originally formulated in 1997 by Duman, Heninger and Nestler, characterizes major depressive disorder as being secondary to aberrant neurogenesis in brain regions that regulate emotion and memory.[1] According to this hypothesis, aberrant neurogenesis is brought about by a (stress induced) lower expression of brain-derived neurotrophic factor (BDNF). In addition, the neurotrophin hypothesis predicts that antidepressants are efficacious because they increase BDNF expression and herewith resolve aberrant neuronal plasticity.[2,3] A large pre-clinical literature, allowing for mechanistic insights, fits very well with these predictions. Taliaz *et al.*,[4] for instance, showed in rats that a reduction of BDNF in the dentate gyrus impairs neurogenesis and induces depressive-like behavior. Human post-mortem studies have indicated similar alternations in the brains of persons who were depressed at the time of dying.[5] Further support for abnormalities in BDNF expression in depressed patients comes from clinical studies. Karege *et al.*[6] as the first found serum BDNF concentrations to be low in depressed patients as compared with healthy controls and lowest in persons with the highest levels of symptom severity. Shimizu *et al.*[7] were the first to show an increase in serum BDNF concentrations in the course of antidepressant treatment.[7]

These findings generated a buzz of research activity and in 2008–2010 the clinical data were summarized in three meta-analyses.[8–10] These meta-analyses, basically including the same 11 studies ($N \sim 968$) confirmed the finding of low serum BDNF concentrations in untreated depressed patients (effect-size (Cohen's d) ~ -1) and normalization of this by antidepressant treatment ($d \sim 1$) while suggesting that these associations were not hampered by between-study heterogeneity or publication bias. Accordingly, the conclusion was: *BDNF may have potential use as biomarker for psychiatric disorders or as a predictor of antidepressant efficacy* (page 527).[9] Since then, the field has seen an abundance of new data on these topics. Important is that this new data entail striking variation in outcomes across studies.[11,12] This, and the abundance of new data, motivated us to update the current state of knowledge by calculating pooled effect-size estimates on differences in serum BDNF concentrations among:

- antidepressant-free depressed patients and healthy controls subjects
- antidepressant-free- and antidepressant-treated depressed patients
- antidepressant-treated depressed patients and healthy controls subjects.

[1]Clinical, Health and Neuropsychology Unit, Leiden University, Leiden, The Netherlands; [2]Leiden Institute for Brain and Cognition, Leiden University Medical Center, Leiden, The Netherlands; [3]Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands; [4]Department of Psychiatry, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; [5]Department of Psychiatry, EMGO Institute and Neuroscience Campus Amsterdam VU, Amsterdam, The Netherlands and [6]University Center for Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. Correspondence: ML Molendijk, Clinical Psychology, Leiden University, Wassenaarseweg 52, Leiden, 2333 AK, The Netherlands.
E-mail: molendijkml@fsw.leidenuniv.nl
Received 28 April 2013; revised 6 July 2013; accepted 18 July 2013; published online 20 August 2013

We chose to focus on serum BDNF concentrations, and not on related parameters such as plasma or whole blood BDNF concentrations, because we wished to avoid an additional source of bias. Besides, in depression research, serum BDNF concentrations are most commonly used.

An additional aim was to compile the data on the putative relation between serum BDNF concentrations and the symptom severity of depression in:

- antidepressant-free depressed patients
- antidepressant-treated depressed patients
- healthy control subjects.

A final aim, made possible by a large amount of studies, was to learn on the potential influence that some relevant moderators might have on the outcomes of our interest.

MATERIALS AND METHODS

We adhered to the guidelines that are recommended by the preferred reporting items for systematic reviews and meta-analyses statement.[13] The literature search, decisions on inclusion, data extraction and quality control were all performed independently by two of the authors (MP or BB and MM).

Search strategy

We searched the PUBMED, Embase and PsychInfo through 1st April 2013 to identify eligible human studies on serum BDNF concentrations in healthy controls, depressed patients or in both. These digital searches were supplemented by backward searches in which the references to the seminal papers of interest were screened[6,7] and by examining the reference sections of the retrieved papers.

Inclusion criteria

We included peer-reviewed human studies that reported data on serum BDNF concentrations in healthy controls, and antidepressant-free and antidepressant-treated depressed patients. Inclusion was independent of clinical (for example, psychiatric comorbidity) and the methodological characteristics of the sample or study (for example, between-subject design versus within-subject design). Non-empirical studies were excluded, as were studies that were not written in English, Dutch, German or Spanish. Overlapping samples were excluded except for the one that reported on the largest number of subjects.

Data extraction

We extracted, as primary outcomes, mean serum BDNF concentrations and s.d. as a function of diagnostic status and antidepressant use and/or indices on the relation between BDNF concentrations and the symptom severity of depression (for example, Pearson's *r*). When BDNF concentrations were assessed at multiple time points, we extracted the data recorded at baseline and at the longest follow-up period.

We also extracted data on mean age, gender distribution, depression severity, antidepressant use (subdivided by selective serotonin reuptake inhibitors, tricyclic antidepressants, selective norepinephrine reuptake inhibitors, and noradrenergic and specific serotonergic antidepressants), duration of antidepressant use and the number of subjects in the study. Where records did not provide sufficient information, corresponding authors were contacted and the required data were requested. In those cases, where nonsignificant results were reported in a paper (for example, $P > 0.05$) and authors did not reply to our request, we assigned the associations an estimated effect-size of zero.

Quality assessment

We used the Newcastle–Ottawa Scale (NOS)[14,15] to assess the quality of the included studies. Overall quality score was defined as the frequency of criteria that were met by the particular study. We excluded NOS items 4 and 7 because these are meaningless in the context of the current paper. Mean-quality score of the included studies was 3.18 (s.d. = 0.14). The agreement between the independent raters was excellent (Cohen's kappa = 0.89, s.e. = 0.03).

Statistical analysis

All calculations were performed using comprehensive meta-analyses 2.0.[16] Random-effects models were applied to calculate pooled Cohen's *d*'s[17] on between-group differences in serum BDNF concentrations. Pooled correlation coefficients were calculated on the relation between serum BDNF concentrations and the symptom severity of depression. All outcomes were weighted using inverse variance methods.[16] Statistical significance of the pooled effect-sizes was assessed using a confidence interval (CI) of 95%. The I^2 measure was used to quantity the amount of between-study heterogeneity and considered to be high when $I^2 > 50\%$.[18,19] Statistical significance of heterogeneity was assessed using the *Q*-statistic.[16]

Through meta-regression analyses, the possible moderating effects of between-study differences on outcomes were evaluated. We considered the number of subjects included in the study, year of publication, mean age, symptom severity of depression of the patient sample, gender distribution and the NOS score as potential moderators for all outcomes of interest. The severity rating scales that were used differed between studies. These instruments use different values to quantify severity (for example, refs. Hamilton *et al.*[20] and Rush *et al.*[21]) that do not necessarily equate to each other. Therefore, we used the validated severity categories (that is, none, mild, moderate, severe and very severe) that can be derived from the continuous scores on each of these instruments as potential moderating variable. The moderation analysis on the difference in serum BDNF concentrations between healthy controls and antidepressant-treated depressed patients in addition included variables coding for the class of antidepressant and the duration of treatment. For the meta-analysis on antidepressant-free and antidepressant-treated depressed patients, the set of moderators was extended with a variable coding for change in depression severity over treatment defined as the percentage of improvement on the depression rating scale that was used.

Publication bias was assessed by inspection of funnel plots and the Egger test.[22] The trim-and-fill procedure, a validated manner to estimate an effect-size after bias has been taken into account,[23,24] was performed in case of publication bias. Power and sample size calculations were performed using G*Power.[25] Stability of our results was evaluated by sensitivity analyses in which each study was excluded from analyses at a time.

RESULTS

Our initial search generated 730 papers of which 55 fulfilled the inclusion criteria for at least one of our meta-analyses.[6,7,11,12,26–76] From these papers, we could extract 124 between-group effect-size estimates and 55 correlation coefficients. For details on the search strategy, we refer to the flowchart (Figure 1). Table 1 lists in which meta-analysis the papers were included and provides demographic and clinical characteristics of the included studies.

Meta-analyses

Random-effects meta-analyses showed that antidepressant-free depressed patients had lower BDNF concentrations as compared with healthy controls ($d = -0.71$, 95% CI = -0.89 to -0.53,

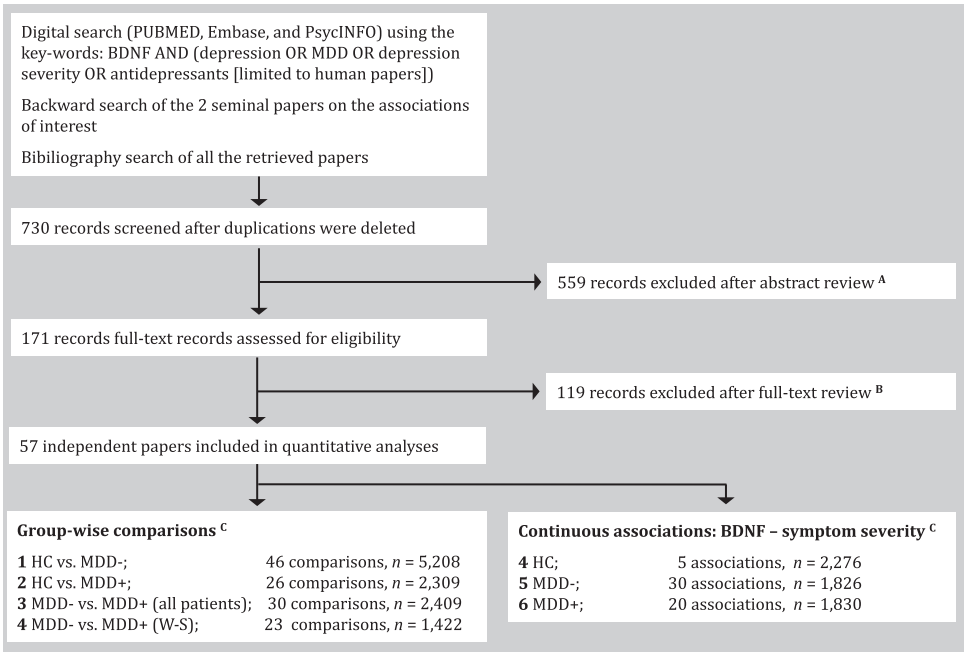

Figure 1. Flowchart of the search strategy and results. BDNF: brain-derived neurotrophic factor; HC, healthy controls; MDD, major depressive disorder. [A]A total of 192 records reported on the BDNF gene, 193 records were reviews, perspectives, comments or hypotheses, 36 records reported on animal data, 14 records were post-mortem studies, 12 records were in *vitro* studies and 111 records did not rapport on BDNF. [B]In all, 2 records reported overlapping data, 3 records reported on the BDNF gene, 64 records reported on plasma BDNF concentrations, 3 records were reviews and 43 records did not reported on serum BDNF concentrations in illnesses other than depression and did not indicate that depression-related assessments were performed. [C]Most of the papers provided input for >1 meta-analytical effect-size. The number of comparisons/associations, therefore, do not add up to 57.

$P<0.0000001$; 46 comparisons, $n=5203$; see Figure 2) and to those of antidepressant-treated depressed patients ($d=-0.56$, 95% CI $=-0.77$ to -0.35, $P<0.00001$, 28 comparisons, $n=4204$). Repeating this latter analysis using only studies that reported pre- and post-treatment BDNF concentrations gave a somewhat higher effect-size estimate ($d=-0.74$, 95% CI $=-1.04$ to -0.45, $P<0.0000001$, 23 comparisons, within-subjects data on 711 patients pre- and post-treatment). Differences in BDNF concentrations among healthy controls and antidepressant-treated depressed patients were not observed ($d=0.07$, $P=0.52$; 24 comparisons, $n=3720$). Forest plots (except Figure 2) are provided as Supplementary materials (Supplementary Figures S1–S3).

A meta-analysis aggregating 30 associations ($n=1807$) on the relation between BDNF concentrations and the symptom severity of depression yielded a statistically significant, negative correlation ($r=-0.19$; 95% CI $=-0.28$ to -0.10, $P<0.00001$) in antidepressant-free depressed patients. There was no evidence for a relation between serum BDNF concentrations and depression severity in antidepressant-treated depressed patients ($r=-0.02$; $P=0.36$, 20 associations, $n=1820$) or in healthy controls ($r=-0.02$; $P=0.41$, 5 associations, $n=2276$). Forest plots are provided as supplement (Supplementary Figures S4–S6).

Between-study heterogeneity and moderation analyses

A large amount of between-study heterogeneity in outcomes was identified in all meta-analyses that yielded significant outcomes ($55\%<I^2<87\%$, for I^2-, Q- and P-values we refer to Table 2).

In a series of meta-regression analyses, we aimed to identify sources of heterogeneity in outcomes. We observed that differences in serum BDNF concentrations among antidepressant-free depressed patients and healthy control subjects could partly be explained by sample size ($r=-0.33$, $R^2=0.11$, $P=0.03$) and by year of publication ($r=-0.30$, $R^2=0.09$, $P=0.04$), with larger samples and more recently reported papers in general reporting smaller between-group differences. In the meta-analysis on changes in serum BDNF concentration over the course of antidepressant treatment, we found that a larger decrease in symptom alleviation was accompanied by a larger increase in BDNF concentrations ($r=-0.48$, $R^2=0.22$, $P=0.01$). Other moderators, including NOS score, were not observed (see Table 3 for all coefficients). Moderation analyses were not performed when between-study heterogeneity was not detected.

Publication bias and power

Visual inspection of the funnel plots suggested that there was evidence for publication bias in all meta-analyses that yielded a significant outcome. Egger's tests confirmed this (t-values in the range 2.5–4.2, P-values all <0.05, see Table 2 for exact values).

Trim-and-fill estimations were used to assess the impact of publication bias. The meta-analysis on differences in BDNF concentrations among healthy controls and untreated depressed patients suggested that nine studies had to be imputed to result in a symetric funnel plot. Imputation led to a smaller, yet significant, effect-size ($d=-0.47$, 95% CI $=-0.64$ to -0.27, $P<0.000001$). The pattern of publication bias was similar in the meta-analyses comparing group differences among antidepressant-free and antidepressant-treated subjects, where five (all data) and four studies (within-subjects data) needed to be imputed to yield a symetric funnel plot. Also here, imputation led to smaller effect-size estimates ($d=-0.54$ and -0.34, respectively, P-values <0.001). Similarly, for the meta-analyses on the continuous association between serum BDNF concentrations and the symptom severity of depression in untreated depressed persons,

Table 1. Summary of study characteristics of included studies (studies are sorted by year and month of publication)

Author	*Meta-analysis*[a]	*Design*[b]	N	*% Female*	*Mean age*	*Patient status*	n[c]	*Severity measure*
Karege *et al.*[6]	(1)(5)	B-S	60	50	37	HC MDD +	30 30	MADRS
Shimizu *et al.*[7]	(1)(2)(3)(5)(6)	Both	83	43	43	HC MDD − MDD +	50 16 17	HAMD
Gervasoni *et al.*[26]	(1)(2)(3)(4)	Both	52	54	40	HC MDD − MDD +	26 26 26	MADRS
Gonul *et al.*[77]	(1)(2)(3)(4)(5)	Both	46	71	36	HC MDD − MDD +	18 28 28	HAMD
Karege *et al.*[28]	(1)(4)	B-S	78	56	34	HC MDD −	35 43	MADRS
Aydemir *et al.*[29]	(1)(2)(3)	Both	20	80	36	HC MDD − MDD +	10 10 10	HAMD
Zanardini *et al.*[30]	(6)	W-S	16	69	56	MDD +	16	HAMD
Lommatzsch *et al.*[31]	(1)(5)	B-S	80	100	28	HC MDD −	62 18	EPDS
Ayedemir *et al.*[29]	(1)(2)(3)	Both	40	100	35	HC MDD − MDD +	20 20 20	HAMD
Bocchi-Chiavetto *et al.*[32]	(6)	W-S	12	70	53	MDD +	12	MADRS
Lang *et al.*[33]	(4)	B-S	24	NK	46	MDD − MDD +	8 16	MADRS
Aydemir *et al.*[34]	(1)	B-S	50	74	33	HC MDD −	26 24	HAMD
Yoshimura *et al.*[35]	(1)(2)(3)(4)	Both	72	65	46	HC MDD − MDD +	30 42 42	HAMD
Ziegenhorn *et al.*[36]	(1)(5)	B-S	465	48	85	HC MDD −	259 91	HAMD
Hellweg *et al.*[37]	(3)	W-S	40	71	51	MDD − MDD +	40 40	HAMD
Okamoto *et al.*[38]	(6)	B-S	18	50	61	MDD +	18	HAMD
Stanek *et al.*[39]	(4)	B-S	34	56	73	HC	34	PRIME-MD
Huang *et al.*[40]	(1)(2)(3)	Both	218	72	33	HC MDD − MDD +	107 111 79	HAMD
Piccini *et al.*[41]	(1)(2)(3)	Both	30	83	42	HC MDD − MDD +	15 15 15	HAMD
Matrisciano *et al.*[42]	(1)(2)(3)	Both	41	51	37	HC MDD − MDD +	20 21 21	HDRS
Basterzi *et al.*[11]	(1)(2)(3)	Both	58	67	33	HC MDD − MDD +	15 43 43	HAMD
Gorgulu *et al.*[43]	(1)(2)(3)	Both	72	69	36	HC MDD − MDD +	31 41 22	HAMD
Grønli *et al.*[44]	(6)	B-S	15	60	70	MDD +	15	HAMD
Umene-Nakano *et al.*[45]	(1)(5)	B-S	40	25	44	HC MDD −	20 20	HAMD
Fernandes *et al.*[46]	(2)(6)	B-S	40	60	42	HC MDD +	30 10	HAMD
Lee *et al.*[47]	(1)	B-S	132	61	74	HC MDD −	98 34	GDS
Ozan *et al.*[48]	(1)	B-S	122	70	34	HC MDD −	56 66	HAMD
Diniz *et al.*[49]	(1)(4)	B-S	71	83	70	HC MDD −	42 29	HAMD
Eker *et al.*[50]	(1)(4)	B-S	47	75	31	HC MDD −	22 25	HAMD
Bocchi-Chiavetto *et al.*[32]	(1)(4)	B-S	84	81	43	HC MDD −	59 25	MADRS
Hu *et al.*[51]	(1)	B-S	84	73	43	HC MDD − MDD −	28 28 28	HAMD
Zhou *et al.*[52]	(1)	B-S	123	NK	NK	HC HC MDD −	30 58 35	HAMD
Su *et al.*[53]	(1)	B-S	52	0	23	HC MDD −	21 31	NK
Rojas *et al.*[54]	(3)	B-S	34	71	42	MDD − MDD +	34 34	HAMD
Yoshimura *et al.*[56]	(3)(4)	W-S	132	60	51	MDD – MDD +	132 132	HAMD
Wolkowitz *et al.*[57]	(1)(2)(3)	B-S	57	36	39	HC MDD − MDD +	28 29 25	HAMD
Kobayakawa *et al.*[58]	(1)	B-S	162	30	65	HC MDD −	81 81	HADS
Terraciano *et al.*[59]	(5)	B-S	2099	62	51	HC MDD −	1661 438	CES-D
Molendijk *et al.*[60]	(1)(2)(3)(4)(5) (6)	B-S	1344	65	42	HC MDD − MDD +	382 541 421	IDS
Toups *et al.*[61]	(6)	B-S	70	80	47	MDD +	70	HAMD
Satomura *et al.*[62]	(2)(4)(5)	B-S	272	63	53	HC MDD +	163 109	HAMD
Sasaki *et al.*[78]	(1)(2)(3)(5)(6)	B-S	52	56	13	HC MDD − MDD +	22 19 11	CDRS-R
Sözeri-Varma *et al.*[63]	(1)(4)	B-S	70	73	37	HC MDD −	40 30	HAMD
Bus *et al.*[64]	(4)	B-S	1230	50	61	HC	1230	BDI
Gedge *et al.*[65]	(5)	W-S	29	69	45	MDD +	29	HAMD
Gazal *et al.*[66]	(1)	B-S	72	100	25	HC MDD −	36 36	BDI
Birkenhäger *et al.*[67]	(5)	W-S	42	43	47	MDD −	42	HAMD
Deuschle *et al.*[68]	(1)(2)(3)(4)	W-S	70	72	52	HC MDD − MDD +	14 56 56	HAMD
Harvey *et al.*[69]	(1)(5)	W-S	200	49	44	HC MDD −	89 111	PHQ-9
Oral *et al.*[70]	(1)(5)	B-S	79	68	27	HC MDD −	40 39	BDI
Karlović *et al.*[71]	(1)	B-S	264	50	46	HC MDD −	142 122	HAMD
Jeon *et al.*[72]	(1)(2)(3)(4)	W-S	155	71	44	HC MDD − MDD +	50 105 105	HAMD
Yoshida *et al.*[73]	(2)(5)	B-S	147	56	38	HC MDD +	78 69	SIGH-D
Elfving *et al.*[12]	(1)(2)	B-S	406	81	46	HC MDD − MDD +	289 117 45	ICD-10
Papakostas *et al.*[76]	(1)	B-S	79	52	36	HC MDD −	43 36	HAMD

Abbreviations: BDI, beck depression inventory; BDNF, brain-derived neurotrophic factor; CDRS-R, children's depression rating scale-revised; CES-D, center for epidemiological studies - depression; HADS, hospital anxiety and depression scale; HAMD, Hamilton depression scale; HC, healthy controls; ICD-10, International classification of disease checklist of symptoms; IDS, inventory of depressive symptoms; MDD −, antidepressant-free MDD; MDD +, antidepressant-treated MDD patients; NK, not known; PHQ-9, patient health questionnaire-9; SIGH-D, structured interview of the Hamilton depression scale. [a]This column indicates in which meta-analysis the study that is indicated in the corresponding row is included: (1) HCs vs MDD −; (2) HCs vs MDD +; (3) MDD − patients vs MDD +; (4–6) regard meta-analyses on continuous associations between serum BDNF concentrations and depression symptom severity scores: (4) in HC's; (5) in MDD −; (6) in MDD +. [b]This column, Design, indicates whether within-subjects data (W-S), a between-subjects data (B–S), or a combination of these types of data (both) is used by the study that is indicated in the corresponding row. [c]Note that the numbers in the column *n* do not add to the numbers as they are given in the column *N*. This is because the numbers in column *n*, in some instances, are counted double (for example, before and after antidepressant treatment in longitudinal designs).

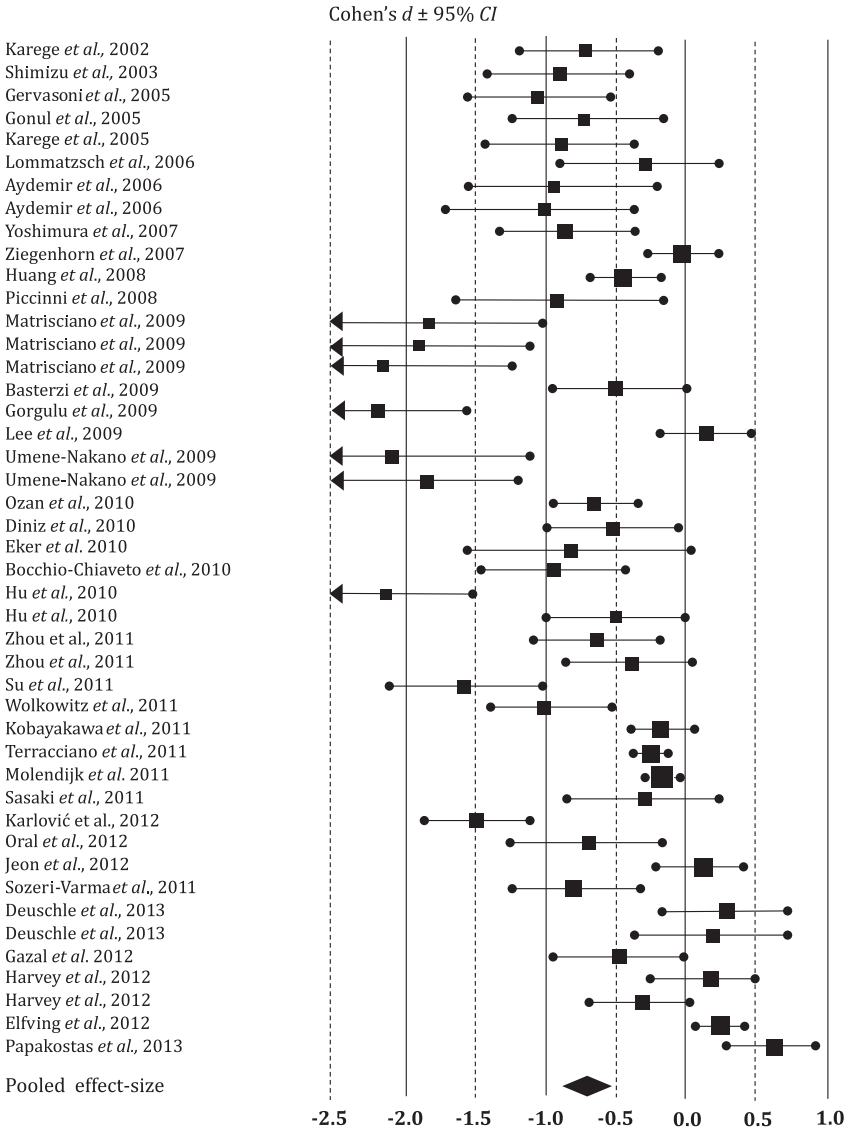

Figure 2. Forrest plot for random-effect meta-analysis on differences in serum BDNF concentrations between healthy control subjects and antidepressant-free depressed patients. The sizes of the squares are proportional to sample size.

Table 2. Statistics on between-study heterogeneity and publication bias for the meta-analysis indicated in the row

	No. of associations/ comparisons	*No. of subjects*			*Heterogeneity*			*Publication bias*	
		HC	*MDD −*	*MDD +*	I^2	Q	P-*value*	*Egger's* t	P-*value*
Group-wise comparisons									
HC vs MDD −	41	2911	2292	NA	86.1%	287.6	<0.0001	4.2	<0.0001
HC vs MDD +	24	2591	NA	1129	84.6%	150.2	<0.0001	1.4	0.16
MDD − vs MDD +	27	NA	2955	1249	84.4%	165.1	<0.0001	2.5	<0.05
MDD − vs MDD + W-S only[a]	23	NA	711	711	83.9%	136.8	<0.0001	2.6	<0.05
Continuous associations									
HC	5	2276	NA	NA	14.8%	4.7	0.32	1.0	0.15
MDD −	29	NA	1807	NA	67.9%	87.2	<0.0001	2.5	<0.05
MDD +	19	NA	NA	1820	18.3%	48.9	0.36	0.6	0.53

Abbreviations: HC, healthy controls; MDD −, antidepressant-free MDD; MDD +, antidepressant-treated MDD patients; NA, not applicable; W-S, within-subjects data.
[a]Here, only associations were included that were derived using a within-subjects designs (that is, treatment studies).

Table 3. Associations (Pearson's correlation coefficients for continuous variables and Spearman's Rho correlation coefficients for categorical variables) between-study characteristics and study effect-size (by meta-analysis)

Group differences	*HC vs MDD −*	*HC vs MDD +*	*MDD − vs MDD +*	*MDD − vs MDD + W-S*
	41 effect-sizes n = *5203*	*24 effect-sizes* n = *3720*	*27 effect-sizes* n = *4204*	*23 effect-sizes* n = *1422*
Gender (percentage female)	0.16	0.11	0.06	0.08
Age (mean, years)	0.13	−0.11	0.08	0.11
Depression severity (categorical, baseline)	−0.17	−0.10	−0.21	−0.07
Percentage SSRI	NA	0.29	−0.35[b]	−0.34
Percentage TCA	NA	−0.21	0.13	0.11
Percentage SNRI	NA	−0.10	0.17	0.17
Percentage NaSSA	NA	−0.14	0.14	0.15
Duration of treatment (mean, weeks)	NA	−0.34	0.04	0.04
Clinical response on treatment	NA	NA	NA	−0.48[a]
Sample size (*n*)	0.33[a]	−0.15	0.25	0.21
Year of publication	0.30[a]	−0.16	0.18	0.18
Study quality (frequency of criteria met)	0.04	0.06	0.35[b]	0.34

Abbreviations: HC, healthy controls; MDD −, antidepressant-free MDD; MDD +, antidepressant-treated MDD patients; NaSSA, noradrenergic and specific serotonergic antidepressants; NK, not known; SNRI, serotonin-norepinephrine reuptake inhibitors; SSRI, selective serotonin reuptake inhibitors; TCA, tricyclic antidepressants; W-S, within-subjects data only (that is, associations were that were derived using a within-subjects design.
Note: Given that there was no evidence for between-study heterogeneity, moderation analysis was not performed in these subgroups. [a]Statistically significant at $P<0.05$. [b]Trend-like finding at $P<0.10$.

the trim-and-fill estimations suggested that five studies had to be imputed to result in a symetric funnel plot pattern. Herewith, the effect-size estimate ($r=-0.07$) was no longer statistically significant. Funnel plots are provided as Supplement (Supplementary Figures S7–S10).

We calculated the numbers of subjects that are needed to detect differences with a power of 0.80 at an α-level of 0.05 (one-sided). Hereto we used the pooled effect-size estimates that were corrected for publication bias. These calculations suggested that 57 subjects in each group would be necessary to reliably detect differences in serum BDNF concentrations between healthy controls and antidepressant-free depressed subjects. For differences in serum BDNF concentrations among antidepressant-free and antidepressant-treated persons, this number would be 108. Based on this, the majority of the included samples was not sufficiently powered (observed median sample size = 36). Sample-size calculations were not performed for continuous associations between serum BDNF concentrations and the symptom severity of depression because these were not statistically significant.

Sensitivity analyses indicated that none of the study findings was unduly driven by the effect of a particular study.

DISCUSSION

Here we confirm, based on a manifold of data as compared with previous meta-analyses,[8–10] that serum BDNF concentrations are low in untreated depressed patients and normalized by antidepressant treatment. The moderate-to-large effect-sizes that we rapport on these differences (random-effects meta-analyses, $d=-0.71$ and -0.56, respectively) are similar to the ones that were reported in the seminal studies[6,7] and in previous meta-analyses.[8–10] These findings are not new. The novelty of our work, instead is that here we highlight a large amount of unexplained between-study heterogeneity in outcomes, underpowered study designs, publication bias that together may call for a critical interpretation of the claim that altered serum BDNF concentrations are related to the illness major depression.

We find a large amount of between-study heterogeneity in outcomes and none of the theoretically relevant variables that we tested (for example, the symptom severity of depression or gender distribution of the sample) was associated with this. Obviously, the heterogeneity may have come from between-sample characteristics that were not tested in our study, such as alcohol consumption and smoking,[79] sleep problems,[80] seasonality[81] or exposure to trauma.[82] Given that depression is a heterogeneous illness,[83] heterogeneity in outcomes may also have come from diversity in clinical characteristics of patient samples. The severity of depression, however, did not explain it. Unfortunately, we did not have the opportunity to test many of the other clinical characteristics because most of the included studies did not report on these variables.

We did find an artificial base for the heterogeneity in outcome. First, a large part of the studies included in our meta-analysis was underpowered. Given that a low level of power increases the false versus true-positive ratio,[84] some overly positive findings may have been among the studies that we included, causing heterogeneity in outcomes. Second, we found that sample size and year of publication were significant predictors of between-study heterogeneity, with larger samples and more recently published findings being associated with smaller between-group differences. This points to publication bias; a particular threat to the validity of a meta-analysis.[85] We found evidence for publication bias in funnel plots.[86,87] Thus, we applied validated trim-and-fill procedures to provide effect-size estimates that can account for this.[23] These yielded attenuated effect-size estimates that were about half as large as those reported in previous meta-analysis[8–10] and of moderate magnitude at best (that is, $d=-0.47$ through -0.34). The often discussed association between serum BDNF concentrations and the symptom severity of depression,[6,65] for which we initially found evidence, even lost its statistical significance after correcting for publication bias and thus likely does not exist. Given that the relevance of a diagnostic biomarker (that is, a variable that is able to distinguish between diagnostic groups)[88] depends on the magnitude of an effect-size (and not on statistical significance *per se*[88,89]) we conclude that serum BDNF concentrations are likely to be of little clinical use (as has been suggested in two earlier, and excellent reviews[90,91]). Complicating this even more is that low serum BDNF concentrations have been reported in persons diagnosed with schizophrenia,[92] bipolar disorder,[93] eating disorders[94] and anxiety[95] indicating that serum BDNF concentrations are not specific enough to differentiate among diagnoses. Multiple-assay methods may serve a role as biomarker better, as recently has been shown.[76]

Although limited in scope with regard to clinical utility, our findings do not dismiss the possibility that abnormalities in BDNF expression reflect the pathophysiological processes that may underlie depressive illnesses.[1,3] It should be noted that the associations that we report on, even when adjusting for publication bias, stand out as being strong when compared with other biological abnormalities in depression, for instance blood markers for immune dysregulation (for example, C-reactive protein and interleukin-6; $d = 0.15$ and 0.25, respectively) or HPA-axis activity (for example, adrenocorticotropin hormone ($d = 0.28$), for a review on these abnormalities see Penninx *et al*.[96]). A difficulty that remains, however, is that we studied peripheral BDNF concentrations. There are indications that BDNF concentrations measured in serum reflect BDNF activity in the brain.[5,97,98] However, it has never been proven that peripheral BDNF concentrations directly reflect or influence the pathophysiology of depression. A complication in understanding this is that other tissues than the brain, including immune-, liver-, smooth muscle- and vascular endothelial cells serve as sources of BDNF.[99,100] The lower peripheral BDNF concentrations in depression and upregulation of this in the course of antidepressant treatment therefore may be an epiphenomenon resulting from an altered BDNF expression (or metabolism) by these peripheral organs. Therefore, the alternations that we rapport on do not necessarily indicate that similar alternations occur at a central level and conclusions with regard to depression related processes in the brain should not be overbearing.

Strengths and limitations

The work presented herein has an obvious strength that it is based on a large amount of data (total $N = 9484$), yielding in general accurate effect-size estimates.[87] Another strength is that through sensitivity- and moderation analyses, we addressed the potential influence of single studies and sources of heterogeneity. Notwithstanding this, our work carries limitations that need to be reflected upon.

Some limitations regard the methods that we used. First, we relied on funnel plot asymmetry and trim-and-fill estimations to assess publication bias. These methods are limited in that one never knows whether asymmetry in a funnel plot is due to publication bias or to unmeasured differences between studies[86] and whether the most extreme effect-sizes are the ones that are left unpublished.[23] Second, in at least some regards the methods that we used were limited with regard to their ability to detect associations. The meta-regression analyses, for instance, may have been underpowered. Besides, *P*-values were not adjusted for multiple comparisons. Also methodologically important is that there may have been noise in our assessment of individual study quality. The NOS scale that we used to this end, although recommended by the Cochrane Collaboration,[15] is not rigorously validated and therefore our quality assessments may have been unreliable.[101] Together, this may have limited our ability to detect true associations or may have led to the detection of associations that in reality do not exist. Finally, our findings are limited in scope in that they cannot be directly generalized to other BDNF parameters such as plasma or whole blood BDNF concentrations because there is no one-to-one relationship among these measures.[32,59,72]

Future work

There are several issues that deserve future research attention. First, our finding that a greater increase in serum BDNF concentrations in the course of antidepressant treatment is associated with greater treatment efficacy may fuel work into the temporal dynamics between BDNF expression and treatment efficacy. It would be interesting if future studies could address early changes in the course of (non-)pharmacological treatment, a notion for which some evidence exists.[102–104] The prediction of how successful a given treatment will be, based on changes in serum BDNF concentrations (that is, *a treatment biomarker*), is a clinically interesting and relevant topic.[88] In our meta-analysis, we did not have the possibility to address this because most of the included studies reported on pre- and post BDNF concentrations only. Another venue for future investigations regards the distinction between the pro- and the mature BDNF variant. The ELISA kits that currently are in use to quantify BDNF are not sensitive enough to make this distinction. Given the proposed opposing effects of these two BDNF variants (proBDNF is believed to induce apoptosis)[2] it would be interesting to study pro/mature BDNF ratios and whether these differ among diagnostic groups. The tools hereto were only recently developed and validated.[73,105]

With regard to future work on peripheral BDNF concentrations, we finally wish to note that analyses would gain credibility if they were controlled for relevant confounding factors and performed using data (preferably within-subject) on a sufficiently large sample ($N \sim 150$, according to our power-analyses).

Concluding remarks

Our meta-analyses (aggregating 179 effect-size estimates; $N = 9484$) initially yielded support for the claim that alternations in serum BDNF concentrations are peripheral manifestations of depression. This is not new. The important contribution of our work is that we clearly show that between-study heterogeneity, underpowered designs and publication bias are at play that give rise to inflated effect-size estimates. Together this suggest that the evidence base for the claim that altered serum BDNF concentrations are peripheral manifestations of depression is slimmer as was initially thought and amidst a lot of noise.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We wish to thank the five reviewers for their helpful comments on our manuscript, the authors who on request kindly provided us with data, and Jeroen Claessen (MSc) for a textual review of an early version of the manuscript. This project is financed with NWO VIDI-grant (Dutch Governmental Scientific Organization, grant number: 016.085.353) awarded to Bernet Elzinga and a KNAW grant (Dutch Governmental Scientific Organization, grant: Programma Akademie Assistent).

AUTHOR CONTRIBUTIONS

All authors participated in interpreting the study outcomes and manuscript redaction and approved the final version of the paper. BE, BP, PS and MM designed the study. BB and MM performed the literature search. MP and MM extracted the data and performed quality assessments. MM conducted the statistical analyses and wrote the paper.

REFERENCES

1 Duman RS, Heninger GR, Nestler EJ. A molecular and cellular theory of depression. *Arch Gen Psychiatry* 1997; **54**: 597–606.
2 Park H, Poo M-M. Neurotrophin regulation of neuronal circuit development and function. *Nat Rev Neurosci* 2013; **14**: 7–23.
3 Duman RS, Monteggia LM. A neurotrophic model for stress-related mood disorders. *Biol Psychiatry* 2006; **59**: 1116–1127.
4 Taliaz D, Stall N, Dar DE, Zangen A. Knockdown of brain-derived neurotrophic factor in specific brain sites precipitates behaviors associated with depression and reduces neurogenesis. *Mol Psychiatry* 2010; **15**: 80–92.
5 Thompson Ray M, Weickert CS, Wyatt E, Webster MJ. Decreased BDNF, trkB-TK + and GAD mRNA expression in the hippocampus of individuals with schizophrenia and mood disorders. *J Psychiatry Neurosci* 2011; **36**: 195–203.

6 Karege F, Perret G, Bondolfi G, Schwald M, Bertschy G, Aubry JM. Decreased serum brain-derived neurotrophic factor levels in major depressed patients. *Psychiatry Res* 2002; **109**: 143–148.
7 Shimizu E, Hashimoto K, Okamura N, Koike K, Komatsu N, Kumakiri C *et al.* Alternations of serum concentrations of brain-derived neurotrophic factor (BDNF) in depressed patients with or without antidepressants. *Biol Psychiatry* 2003; **54**: 70–75.
8 Brunoni AR, Lopes M, Fregni F. A systematic review and meta-analysis of clinical studies on major depression and BDNF levels: implications for the role of neuroplasticity in depression. *Int J Neuropsychopharmacol* 2008; **11**: 1169–1180.
9 Sen S, Duman R, Sanacora G. Serum brain-derived neurotrophic factor, depression, and antidepressant medications: meta-analyses and implications. *Biol Psychiatry* 2008; **64**: 527–532.
10 Bocchio-Chiavetto L, Bagnardi V, Zanardini R, Molteni R, Nielsen MG, Placentino A *et al.* Serum and plasma BDNF levels in major depression: a replication study and meta-analyses. *W J Biol Psychiatry* 2010; **11**: 763–773.
11 Basterzi AD, Yazici K, Aslan E, Delialioglu N, Tasdelen B, Acar ST *et al.* Effects of fluoxetine and venlafaxine on serum brain-derived neurotrophic factor concentrations in depressed patients. *Prog Neuro-Psychopharmacol Biol Psychiatry* 2009; **33**: 281–285.
12 Elfving B, Buttenschøn HN, Foldager L, Poulsen PHP, Andersen JH, Grynderup MB *et al.* Depression, the val^{66}met polymorphism, age, and gender influence the serum BDNF level. *J Psychiatric Res* 2012; **46**: 1118–1125.
13 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Ann Intern Med* 2009; **151**: 264–269.
14 Wells GA, Shea B, O'Connel D. The Newcastle-Ottawa Scale (NOS) for assessing the quality non-randomised studies in meta-analyses. Ottawa Healthcare Institute. http://www.ohri.ca, accessed 7 February 2013.
15 The Cochrane Collaboration Handbookhttp://www.cochrane.org/, accessed 29 June 2013.
16 Borenstein M, Hedges LV, Higgins JPT. *Introduction to Meta-Analyses*. Wiley: Chicester, NH, USA, 2009.
17 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.
18 Mosteller F, Colditz GA. Understanding research synthesis (meta-analysis). *Annu Rev Public Health* 1996; **17**: 1–23.
19 Higgins JPT, Thompson SG. Quantifying heterogeneity in meta-analysis. *Statist Med* 2002; **21**: 1539–155.
20 Hamilton M. A rating scale for depression. *J Neurol Neurosug Psychiatry* 1960; **23**: 56–62.
21 Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The inventory of depressive symptomatology (IDS): psychometric properties. *Psychol Med* 1996; **26**: 477–486.
22 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629–634.
23 Duval S, Tweedie RL. A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *J Am Stat Soc* 2000; **95**: 89–98.
24 Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statist Med* 2007; **26**: 4544–4562.
25 Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Beh Res Methods* 2009; **41**: 1149–1160.
26 Gervasoni N, Aubry JM, Bondolfi G, Osiek C, Schwald M, Bertschy G *et al.* Partial normalization of serum brain-derived neurotrophic factor in remitted patients after a major depressive episode. *Neuropsychobiol* 2005; **51**: 234–238.
27 Gönül AS, Kitis O, Eker MC, Eker OD, Ozan E, Coburn K. Association of the brain-derived neurotrophic factor val^{66}met polymorphism with hippocampus volumes in drug-free depressed patients. *W J Biol Psychiatry* 2011; **12**: 110–118.
28 Karege F, Bondolfi G, Gervasoni N, Schwald M, Aubry J-M, Bertschy G. Low brain-derived neurotrophic factor levels in serum of depressed patients probably result from lower platelet BDNF release unrelated to platelet reactivity. *Biol Psychiatry* 2005; **57**: 1068–1072.
29 Aydemir C, Yalcin ES, Aksaray S, Kisa C, Yildirim SG, Uzbay T *et al.* Brain-derived neurotrophic factor (BDNF) changes in the serum of depressed women. *Prog Neuropsychopharmacol Biol Psychiatry* 2006; **30**: 1256–1260.
30 Zanardini R, Gazzoli A, Ventriglia M, Perez J, Bignotti S, Rossini PM *et al.* Effect of repetitive transcranial magnetic stimulation on serum brain-derived neurotrophic factor in drug resistant depressed patients. *J Affect Disord* 2006; **91**: 83–86.
31 Lommatzsch M, Hornych K, Zingler C, Schuff-Werner P, Hoppner J, Virchow JC *et al.* Maternal serum concentrations of BDNF and depression in the perinatal period. *Psychoendocrinol* 2006; **31**: 388–394.
32 Bocchi-Chiavetto L, Zanardini R, Bortolomasi M, Abate M, Segala M, Giacopuzzi M *et al.* Electroconvulsive therapy (ECT) increases serum brain-derived neurotrophic factor in drug resistant depressed patients. *Eur Neuropsychopharmacol* 2006; **16**: 56–59.
33 Lang UE, Bajbouj M, Gallinat J, Hellweg R. Brain-derived neurotrophic factor serum concentrations in depressive patients during vagus nerve stimulation and repetetive transcranial magnetic stimulation. *Psychopharmacol* 2006; **187**: 56–59.
34 Aydemir O, Deveci A, Taskin OE, Taneli F, Esen-Danaci A. Serum brain-derived neurotrophic factor level in dysthymia: a comparative study with major depressive disorder. *Prog Neuropsychopharmacol Biol Psychiatry* 2007; **31**: 1023–1026.
35 Yoshimura R, Mitoma M, Sugita A, Hori H, Okamoto T, Umene W *et al.* Effects of paroxetine or milnacipran on serum brain-derived neurotrophic factor in depressed patients. *Prog Neuropsychopharmacol Biol Psychiatry* 2007; **31**: 1034–1037.
36 Ziegenhorn AA, Schulte-Hebruggen O, Danker-Hopfe H, Malbranc M, Hartung H-D, Anders D. Serum neurotrophins—a study on the time course and influencing factors in a large old sample. *Neurobiol Aging* 2007; **28**: 1436–1445.
37 Hellweg R, Ziegenhorn A, Heuser I, Deuschle M. Serum concentrations of nerve growth factor and brain-derived neurotrophic factor in depressed patients before and after antidepressant treatment. *Pharmacopsychiatry* 2008; **41**: 66–71.
38 Okamoto T, Yoshimura R, Ikenouchi-Sugita A, Hori H, Umene-Nakano W, Inoue Y *et al.* Efficacy of electroconvulsive therapy is associated with changing blood levels of homovanilic acid and brain-derived neurotrophic factor in refractory depressed patients: a pilot study. *Prog Neuropsychopharmacol Biol Psychiatry* 2008; **32**: 1185–1190.
39 Stanek K, Gunstad J, Leahey T, Glickman E, Alexander T, Spitznagel MB *et al.* Serum brain-derived neurotrophic factor is associated with reduced appetite in healthy older adults. *J Health Aging* 2008; **12**: 183–185.
40 Huang TL, Lee CT, Liu YL. Serum brain-derived neurotrophic factor levels in patients with major depression: effects of antidepressants. *J Psychiatr Res* 2008; **42**: 521–525.
41 Piccinni A, Marazziti D, Catena M, Domenici L, Del Debbio A, Bianchi C *et al.* Plasma and serum brain-derived neurotrophic factor (BDNF) in depressed patients during 1 year of antidepressant treatment. *J Affect Disord* 2008; **105**: 279–283.
42 Matrisciano F, Bonaccorso S, Ricciardi A, Scaccianoce S, Panaccione I, Wang L *et al.* Changes in BDNF serum concentrations in patients with major depression disorder (MDD) after 6 months treatment with sertraline, ecitalopram, or venlafaxine. *J Psych Res* 2009; **43**: 247–254.
43 Gorgulu Y, Caliyurt O. Rapid antidepressant effect of sleep deprivation therapy correlates with serum BDNF changes in major depression. *Brain Res Bull* 2009; **80**: 158–162.
44 Grønli O, Stensland GØ, Wynn R, Olstad R. Neurotrophic factors in serum following ECT: a pilot study. *World J Biol Psychiatry* 2009; **10**: 295–301.
45 Umene-Nakano W, Yoshimura R, Ikenouchi-Sugita A, Hori H, Hayashi K, Ueda N *et al.* Serum levels of brain-derived neurotrophic factor in comorbidity of depression and alcohol dependence. *Hum Psychopharmacol Clin Exp* 2009; **24**: 409–413.
46 Fernandes B, Gama CS, Massuda R, Torres M, Camargo D, Kunz M *et al.* Serum brain-derived neurotrophic factor is not associated with response to electroconvulsive therapy: a pilot in drug resistant depressed patients. *Neurosci Lett* 2009; **453**: 195–198.
47 Lee JG, Shin BS, You YS, Kim JE, Yoon SW, Jeon DW *et al.* Decreased serum brain-derived neurotrophic factor levels in elderly Korean with dementia. *Psychiatry Invest* 2009; **6**: 299–305.
48 Ozan E, Okur H, Eker C, Eker DE, Gönül AS, Akarsu N. The effect of depression, BDNF gene val^{66}met polymorphism and gender on serum BDNF levels. *Brain Res Bull* 2010; **81**: 61–65.
49 Diniz BS, Teixeira AL, Talib LL, Mendoca VA, Gattaz WF, Forlenza OV. Serum brain-derived neurotrophic factor level is reduced in antidepressant-free patients with late-life depression. *W J Biol Psychiatry* 2010; **11**: 550–555.
50 Eker C, Kitis O, Taneli F, Eker DE, Ozan E, Yucel K *et al.* Correlation of serum BDNF levels with hippocampal volumes in first episode, medication-free depressed patients. *Eur Arch Psychiatry Clin Neurosci* 2010; **260**: 527–533.
51 Hu Y, Yu X, Yang F, Si T, Wang W, Tan Y *et al.* The brain-derived neurotrophic factor is associated with the therapeutic efficacy of modified electroconvulsive therapy in Chinese patient with depression. *J ECT* 2010; **26**: 121–125.
52 Zhou Z, Lu T, Xu G, Yue X, Zhu W, Ma M *et al.* Decreased serum brain-derived neurotrophic factor (BDNF) is associated with post-stroke depression but not with BDNF gene val^{66}met polymorphism. *Clin Chem Lab Med* 2011; **49**: 185–189.
53 Su SC, Sun MT, Wen MJ, Lin CJ, Chen YC, Hung YJ. Brain-derived neurotrophic factor, adiponectin, and proinflammatory markers in various subtypes of depression in young men. *Int J Psychiatry Med* 2011; **42**: 211–226.
54 Rojas PS, Fritsch R, Rojas RA, Jara P, Fiedler JL. Serum brain-derived neurotrophic factor and glucocorticoid receptor levels in lymphocytes as markers of

antidepressant response in major depressive patients: a pilot study. *Psychiatry Res* 2011; **189**: 239–245.
55 Gorgulu Y, Caliyurt O. Rapid antidepressant effects of sleep deprivation therapy correlates with serum BDNF changes in major depression. *Brain Res Bull* 2009; **80**: 158–162.
56 Yoshimura R, Kishi T, Suzuki A, Umene-Nakano W, Ikenouchi-Sugita A, Hori H *et al.* The brain-derived neurotrophic factor (BDNF) polymorphism val^{66}met is associated with neither serum BDNF level nor response to selective serotonin reuptake inhibitors in depressed Japanese patients. *Prog Neuropsychopharmacol Biol Psychiatry* 2011; **35**: 1022–1025.
57 Wolkowitz OM, Wolf J, Shelly W, Rosser R, Burke HM, Lerner GK *et al.* Serum BDNF levels before treatment predict SSRI response in depression. *Prog Neuropsychopharmacol Biol Psychiatry* 2011; **15**: 1623–1630.
58 Kobayakawa M, Inagaki M, Fujimori M, Hamazaki K, Hamazaki T, Akechi T *et al.* Serum brain-derived neurotrophic factor and antidepressant-naive major depression after lung cancer diagnosis. *Jpn J Clin Oncol* 2011; **4**: 1233–1237.
59 Terracciano A, Lobina M, Piras MG, Cannas A, Meirelles O, Mulas A *et al.* Lower serum BDNF associated with higher neuroticism and depressive symptoms. *Psychosom Med* 2011; **74**: 638–642.
60 Molendijk ML, Bus BAA, Spinhoven P, Penninx BJWH, Kenis G, Prickaerts J *et al.* Serum concentrations of brain-derived neurotrophic factor in major depressive disorder: state and trait issues, clinical features, and pharmacological treatment. *Mol Psychiatry* 2011; **16**: 1088–1095.
61 Toups MSP, Greer TL, Kurian BT, Grannemann BD, Carmody TJ, Huebinger R *et al.* Effects of serum brain-derived neurotrophic factor on exercise augmentation treatment of depression. *J Psychiatric Res* 2011; **45**: 1301–1306.
62 Satomura E, Baba H, Nakano Y, Maeshima H, Suzuki T, Arai H. Correlations between brain-derived neurotrophic factor and clinical symptoms inmedicated patients with major depression. *J Affect Disord* 2011; **135**: 332–335.
63 Sözeri-Varma G, Enli Y, Toker-Uğurlu T, Alaçam H, Kalkan-Oğuzhanoğlu N. Decreased serum BDNF levels in major depressive patients. *Neurol Psychiatry Brain Res* 2011; **17**: 84–88.
64 Bus BAA, Tendolkar I, Franke B, de Graaf J, Den Heijer M, Buitelaar JK *et al.* Serum brain-derived neurotrophic factor: determinants and relationship with depressive symptoms in a community population of middle-aged and elderly people. *W J Biol Psychiatry* 2012; **13**: 39–47.
65 Gedge L, Beaudoin A, Lazowski L, Toit R, Jokic R, Milev R. Effects of electroconvulsive therapy and repetetive transcranial magnetic stimulation on serum brain-derived neurotrophic factor levels in patients with depression. *Front Psychiatry* 2012; **3**: 1–8.
66 Gazal M, Motta LS, Wiener CD, Fernandes JC, Quevedo LA, Jansen K *et al.* Brain-derived neurotrophic factor in post-partum depressive mothers. *Neurochem Res* 2012; **37**: 583–587.
67 Birkenhäger TK, Geldermans S, Van den Broek WW, van Beveren N, Fekkes D. Serum brain-derived neurotrophic factor level in relation to illness severity and episode duration in patients with major depression. *J Psychiatr Res* 2012; **46**: 285–289.
68 Deuschle M, Gilles M, Scharnholz B, Lederbogen F, Lang U, Hellweg R. Changes of serum concentrations of brain-derived neurotrophic factor (BDNF) during treatment with venlafaxine and mirtazapine: role of medication and response to treatment. *Pharmacopsychiatry* 2013; **46**: 54–58.
69 Harvey BH, Hamer M, Louw R, van der Westhuizen FH, Malan L. Metabolic and glutathione redox markers associated with brain-derived neurotrophic factor in depressed African men and women: evidence for counter-regulation. *Neuropsychobiol* 2012; **67**: 33–40.
70 Oral E, Canpolat S, Yildirim S, Gulec M, Aliyev E, Aydin N. *et al.* Cognitive functions and serum levels of brain-derived neurotrophic factor in patients with major depressive disorder. *Brain Res Bull* 2012; **88**: 454–459.
71 Karlović D, Serretti A, Jevtović S, Vrkić N, Šeric V, Peleš AM. Diagnostic accuracy of serum brain derived neurotrophic factor concentration in antidepressant naïve patients with first major depression episode. *J Psychiatric Res* 2012; **47**: 162–167.
72 Jeon HJ, Kang E-S, Lee EH, Jeong E-H, Jeon JR, Mischoulon D *et al.* Childhood trauma and platelet brain-derived neurotrophic factor (BDNF) after a three month follow-up in patients with major depressive disorder. *J Psychiatric Res* 2012; **46**: 966–972.
73 Yoshida T, Ishikawa M, Niitsu T, Nakazato M, Watanabe H, Shiraishi T *et al.* Decreased serum levels of mature brain-derived neurotrophic factor, but not Its precursor proBDNF, in patients with major depressive disorder. *PLoS ONE* 2012; **7**: e42676.
74 Gazal M, Motta LS, Wiener CD, Fernandes JC, Quevedo LA, Jansen K *et al.* Brain-derived neurotrophic factor in post-partum depressive mothers. *Neurochem Res* 2012; **37**: 583–587.
75 Oral E, Canpolat S, Yildirim S, Gulec M, Aliyev E, Aydin N. Cognitive functions and serum brain derived neurotrophic factor in patients with major depressive disorder. *Brain Res Bull* 2012; **88**: 454–459.
76 Papakostas GI, Shelton RC, Kinrys G, Henry ME, Bakow BR, Lipkin SH *et al.* Assessment of multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study. *Mol Psychiatry* 2013; **18**: 332–339.
77 Gonul AS, Akdeniz F, Taneli F, Donat O, Eker Ç, Vahip S. Effect of treatment on serum brain-derived neurotrophic factor levels in depressed patients. *Eur Arch Psychiatry Clin Neurosci* 2005; **255**: 381–386.
78 Sasaki T, Niitsu T, Hashimoto T, Kanahara N, Shiina A, Hasegawa T *et al.* Decreased levels of between brain-derived neurotrophic factor in male pediatric patients with depression. *Open Clin Chem J* 2011; **4**: 28–33.
79 Bus BAA, Molendijk ML, Penninx BJWH, Buitelaar JK, Kenis G, Prickaerts J *et al.* Determinants of serum brain-derived neurotrophic factor. *Psychoneuroendocrinol* 2011; **36**: 228–239.
80 Giese M, Unternaher E, Huttig H, Beck J, Brand S, Calabrese P *et al.* BDNF: an indicator of insomnia? *Mol Psychiatry* advance online publication, 12 February 2013; doi:10.1038/mp.2013.10.
81 Molendijk ML, Haffmans JPM, Bus BAA, Spinhoven P, Penninx BWJH, Prickaerts J *et al.* BDNF concentrations show strong seasonal variation and correlations with the amount of ambient sunlight. *PLoS One* 2012; **7**: e48046.
82 Elzinga BM, Molendijk ML, Oude Voshaar RC, Bus AA, Prickaerts J, Spinhoven P *et al.* The impact of childhood abuse and recent stress on serum brain-derived neurotrophic factor and the moderating role of BDNF val^{66}met. *Psychopharmacol* 2011; **214**: 319–328.
83 Rush AJ. The varied clinical presentations of major depressive disorders. *J Clin Psychiatry* 2007; **68**: S4–S10.
84 Sterne JAC, Davey Smith G. Sifting the evidence – what's wrong with significance tests? *BMJ* 2001; **322**: 226–231.
85 Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990; **263**: 1385–1359.
86 Munafo MR, Flint J. Meta-analysis of genetic association studies. *Trends Gen* 2004; **20**: 439–444.
87 Ioannidis JPA. Why most published research findings are false. *PLoS MED* 2005; **2**: e124 89.
88 Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry* 2012; **17**: 1174–1179.
89 Schmidt HD, Shelton RC, Duman RS. Functional biomarkers of depression: diagnosis, treatment, and pathophysiology. *Neuropsychopharmacol* 2011; **36**: 2375–2394.
90 Groves JO. Is it time to reassess the BDNF hypothesis of depression? *Mol Psychiatry* 2007; **12**: 1079–1088.
91 Gass P, Hellweg R. Peripheral brain-derived neurotrophic factor (BDNF) levels as a biomarker for affective disorders. *Int J Neuropsychopharmacol* 2010; **13**: 1–4.
92 Green MJ, Matheson SL, Sheperd A, Weickert CS, Carr VJ. Brain-derived neurotrophic factor levels in schizophrenia: a systematic review with meta-analysis. *Mol Psychiatry* 2011; **16**: 960–972.
93 Fernandes BS, Gama CS, Cereser KM, Yatham LN, Fries GR, Colpo G *et al.* Brain-derived neurotrophic factor as a state marker of mood episodes in bipolar disorders: a systematic review and meta-regression analysis. *J Psychiatric Res* 2011; **45**: 995–1004.
94 Monteleone P, Fabrazzo M, Martiadis V, Serritella C, Pannuto M, Maj M. Circulating brain-derived neurotrophic factor is decreased in women with anorexia and bulimia nervosa but not in women with binge eating disorder: relationships to co-morbid depression, psychopathology and hormonal variables. *Psychol Med* 2005; **35**: 897–905.
95 Molendijk ML, Bus BAA, Spinhoven P, Penninx BWJH, Prickaerts J, Oude Voshaar RC *et al.* Gender specific associations of serum BDNF concentrations in anxiety. *W J Biol Psychiatry* 2012; **13**: 535–543.
96 Penninx BWJH, Milaneschi Y, Lamers F, Vogelzangs N. Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. *BMC Med* 2013; **11**: 129.
97 Dawood T, Anderson J, Barton D, Lambert E, Esler M, Hotchkin E *et al.* Reduced overflow of BDNF from the brain is linked with suicide risk in depressive illness. *Mol Psychiatry* 2007; **12**: 981–983.
98 Klein AB, Williamson R, Santini MA, Clemmensen C, Ettrup A, Rios M *et al.* Blood BDNF concentrations reflect brain-tissue BDNF concentrations across species. *Int J Neuropsychopharmacol* 2011; **14**: 347–353.
99 Cassiman D, Denef C, Desmet VJ, Roskams T. Human and rat hepatic stellate cells express neurotrophins and neurotrophin receptors. *Hepatology* 2001; **33**: 148–158.
100 Karege F, Schwald M, Cisse M. Postnatal developmental profile of brain-derived neurotrophic factor in rat brain and platelets. *Neurosci Lett* 2002; **328**: 261–264.
101 Sanderson S, Tatt ID, Higgings JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007; **36**: 666–676.

102 aan het Rot M, Zarate CA, Charney DS, Mathew SJ. Ketamine for depression: where do wego from here? *Biol Psychiatry* 2012; **72**: 537–547.
103 Machado-Vieira R, Yuan P, Brutsche N, Diaz-Granados N, Luckenbaugh D, Manji HK *et al*. Brain-derived neurotrophic factor and initial antidepressant response to an N-methyl-D- aspartate antagonist. *J Clin Psychiatry* 2009; **70**: 1662–1666.
104 Lang UE, Bajbouj M, Gallinat J, Hellweg R. Brain-derived neurotrophic factor serumconcentrations in depressive patients during vagus nerve stimulation and repetetive transcranial magnetic stimulation. *Psychopharmacol* 2006; **187**: 56–59.
105 Yoshida T, Ishikawa M, Masaomi I, Hashimoto K. Serum levels of mature brain-derived neurotrophic factor (BDNF) and its precursor proBDNF in healthy subjects. *Open ClinChem* 2012; **5**: 7–12.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)

Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: a cross-sectional analysis

Jonathan D Rohrer, Jennifer M Nicholas, David M Cash, John van Swieten, Elise Dopper, Lize Jiskoot, Rick van Minkelen, Serge A Rombouts, M Jorge Cardoso, Shona Clegg, Miklos Espak, Simon Mead, David L Thomas, Enrico De Vita, Mario Masellis, Sandra E Black, Morris Freedman, Ron Keren, Bradley J MacIntosh, Ekaterina Rogaeva, David Tang-Wai, Maria Carmela Tartaglia, Robert Laforce Jr, Fabrizio Tagliavini, Pietro Tiraboschi, Veronica Redaelli, Sara Prioni, Marina Grisoli, Barbara Borroni, Alessandro Padovani, Daniela Galimberti, Elio Scarpini, Andrea Arighi, Giorgio Fumagalli, James B Rowe, Ian Coyle-Gilchrist, Caroline Graff, Marie Fallström, Vesna Jelic, Anne Kinhult Ståhlbom, Christin Andersson, Håkan Thonberg, Lena Lilius, Giovanni B Frisoni, Giuliano Binetti, Michela Pievani, Martina Bocchetta, Luisa Benussi, Roberta Ghidoni, Elizabeth Finger, Sandro Sorbi, Benedetta Nacmias, Gemma Lombardi, Cristina Polito, Jason D Warren, Sebastien Ourselin, Nick C Fox, Martin N Rossor

Summary

Background Frontotemporal dementia is a highly heritable neurodegenerative disorder. In about a third of patients, the disease is caused by autosomal dominant genetic mutations usually in one of three genes: progranulin (*GRN*), microtubule-associated protein tau (*MAPT*), or chromosome 9 open reading frame 72 (*C9orf72*). Findings from studies of other genetic dementias have shown neuroimaging and cognitive changes before symptoms onset, and we aimed to identify whether such changes could be shown in frontotemporal dementia.

Methods We recruited participants to this multicentre study who either were known carriers of a pathogenic mutation in *GRN*, *MAPT*, or *C9orf72*, or were at risk of carrying a mutation because a first-degree relative was a known symptomatic carrier. We calculated time to expected onset as the difference between age at assessment and mean age at onset within the family. Participants underwent a standardised clinical assessment and neuropsychological battery. We did MRI and generated cortical and subcortical volumes using a parcellation of the volumetric T1-weighted scan. We used linear mixed-effects models to examine whether the association of neuropsychology and imaging measures with time to expected onset of symptoms differed between mutation carriers and non-carriers.

Findings Between Jan 30, 2012, and Sept 15, 2013, we recruited participants from 11 research sites in the UK, Italy, the Netherlands, Sweden, and Canada. We analysed data from 220 participants: 118 mutation carriers (40 symptomatic and 78 asymptomatic) and 102 non-carriers. For neuropsychology measures, we noted the earliest significant differences between mutation carriers and non-carriers 5 years before expected onset, when differences were significant for all measures except for tests of immediate recall and verbal fluency. We noted the largest *Z* score differences between carriers and non-carriers 5 years before expected onset in tests of naming (Boston Naming Test –0·7; SE 0·3) and executive function (Trail Making Test Part B, Digit Span backwards, and Digit Symbol Task, all –0·5, SE 0·2). For imaging measures, we noted differences earliest for the insula (at 10 years before expected symptom onset, mean volume as a percentage of total intracranial volume was 0·80% in mutation carriers and 0·84% in non-carriers; difference –0·04, SE 0·02) followed by the temporal lobe (at 10 years before expected symptom onset, mean volume as a percentage of total intracranial volume 8·1% in mutation carriers and 8·3% in non-carriers; difference –0·2, SE 0·1).

Interpretation Structural imaging and cognitive changes can be identified 5–10 years before expected onset of symptoms in asymptomatic adults at risk of genetic frontotemporal dementia. These findings could help to define biomarkers that can stage presymptomatic disease and track disease progression, which will be important for future therapeutic trials.

Funding Centres of Excellence in Neurodegeneration.

Lancet Neurol 2015; 14: 253–62

Published Online February 4, 2015 http://dx.doi.org/10.1016/S1474-4422(14)70324-2

This online publication has been corrected. The corrected version first appeared at thelancet.com/neurology on November 9, 2015

See Comment page 236

Dementia Research Centre (J D Rohrer PhD, J M Nicholas PhD, D M Cash PhD, M J Cardoso PhD, S Clegg BSc, M Espak PhD, D L Thomas PhD, Prof J D Warren PhD, Prof S Ourselin PhD, Prof N C Fox MD, Prof M N Rossor MD), **and Medical Research Council Prion Unit** (S Mead PhD), **Department of Neurodegenerative Disease, University College London Institute of Neurology, and Centre for Medical Image Computing** (M J Cardoso, M Espak, Prof S Ourselin, D M Cash), **and Neuroradiological Academic Unit, Department of Brain Repair and Rehabilitation** (D L Thomas, E De Vita PhD), **University College London, London, UK; Lysholm Department of Neuroradiology, National Hospital for Neurology and Neurosurgery, London, UK** (E De Vita); **Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK** (J M Nicholas); **Department of Neurology** (Prof J van Swieten MD, E Dopper MD, L Jiskoot MSc),

Introduction

Frontotemporal dementia is a neurodegenerative disorder characterised by focal neuronal loss in the frontal and temporal lobes.[1] It is a common cause of early-onset dementia, but can also present in old age and has an estimated prevalence of between 15 and 22 per 100 000 individuals in the population.[2] It presents clinically with either behavioural symptoms (behavioural variant

and Department of Clinical Genetics (R van Minkelen PhD), Erasmus Medical Center, Rotterdam, Netherlands; Institute of Psychology, Leiden University, and Department of Radiology, Leiden University Medical Center, Leiden, Netherlands (Prof S A Rombouts PhD); LC Campbell Cognitive Neurology Research Unit, Department of Medicine, Division of Neurology, Sunnybrook Health Sciences Centre, Sunnybrook Research Institute (M Masellis MD, Prof S E Black MD, B J MacIntosh PhD), Department of Medicine, Division of Neurology, Baycrest, Mount Sinai Hospital (Prof M Freedman MD), and Tanz Centre for Research in Neurodegenerative Diseases (E Rogaeva PhD, M Carmela Tartaglia MD), University of Toronto, Toronto, ON, Canada; Rotman Research Institute, Baycrest, Toronto, ON, Canada (Prof M Freedman); University Health Network Memory Clinic, Toronto Western Hospital, Toronto, ON, Canada (R Keren MD, D Tang-Wai MD); Clinique Interdisciplinaire de Mémoire, Département des Sciences Neurologiques, Hôpital de l'Enfant-Jésus, and Faculté de Médecine, Université Laval, QC, Canada (R Laforce Jr MD); Fondazione Istituto di Ricovero e Cura a Carattere Scientifico Istituto Neurologico Carlo Besta, Milano, Italy (F Tagliavini MD, P Tiraboschi MD, V Redaelli MD, S Prioni MSc, M Grisoli MD); Neurology Unit, Department of Medical and Experimental Sciences (B Borroni MD, Prof A Padovani MD), and Department of Molecular and Translational Medicine (M Bocchetta MSc), University of Brescia, Brescia, Italy; Neurology Unit, Department of Physiopathology and Transplantation, University of Milan, Fondazione Cà Granda, Istituto di Ricovero e Cura a Carattere Scientifico Ospedale Policlinico, Milan, Italy (D Galimberti PhD, E Scarpini MD, A Arighi MD, G Fumagalli MD); Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK (J B Rowe PhD, I Coyle-Gilchrist MB);

frontotemporal dementia) or language disturbance (primary progressive aphasia), but patients can also develop symptoms of motor neuron disease, progressive supranuclear palsy, or corticobasal syndrome.[1] It is highly heritable, with an autosomal dominant family history reported in around a third of people with the disease.[3] Mutations in three genes are proven major causes of genetic frontotemporal dementia: microtubule-associated protein tau (*MAPT*), progranulin (*GRN*), and chromosome 9 open reading frame 72 (*C9orf72*).[4] Frequencies of mutations of these three genes vary by geography, but together they account for 10–20% of all cases of frontotemporal dementia.[4]

The study of autosomal dominant frontotemporal dementia in its presymptomatic period provides a window into the earliest stages of the disease process.[5] Evidence from familial Alzheimer's disease and Huntington's disease shows that changes in some biomarkers occur many years before symptom onset,[6–8] suggesting that the ideal time to treat neurodegenerative disease could be before clinical presentation, at a point when the minimum of irreversible neuronal loss has occurred and cognitive function is still preserved. To optimise therapeutic opportunities, biomarkers of frontotemporal dementia are therefore needed that signify disease onset and can measure changes in disease trajectory in the presymptomatic period. Furthermore, biomarkers that allow accurate staging of the disease process will be important to identify individuals most suitable for particular trials, to reduce heterogeneity, and increase the statistical power.

Few studies of mutation carriers at risk of frontotemporal dementia have been done, and investigators of these studies have reported inconsistent findings (appendix).[9–26] Although findings from some studies have shown presymptomatic changes in neuropsychometric testing near to disease onset,[9,11,15–17,22] others have not shown any changes.[13,19,21,23–25] Similarly, findings from a few case studies[9,11,17] and small case series[12,13,18] have shown evidence of grey matter volume loss before symptoms onset with structural MRI, but other studies have reported no abnormalities.[19–22] In this study, we compared clinical, behavioural, and structural imaging measures between mutation carriers and non-carriers in a large international cohort of families with autosomal dominant frontotemporal dementia. Our hypothesis was that we would see presymptomatic changes in structural imaging measures initially and then behavioural and cognitive measures before onset of symptoms.

Methods

Participants

The Genetic Frontotemporal dementia Initiative (GENFI) consists of 11 research sites, in the UK, Italy, the Netherlands, Sweden, and Canada. We recruited participants who were either known carriers of a pathogenic mutation in *MAPT*, *GRN*, or *C9orf72*, or at risk of carrying a mutation because a first-degree relative was a known symptomatic carrier. We genotyped all participants at their local site, with a pathogenic expansion in *C9orf72* being defined as the presence of greater than 30 repeats. We enrolled 220 participants between Jan 30, 2012, and Sept 15, 2013. Local ethics committees at each site approved the study and all participants provided written informed consent at enrolment.

Procedures

Participants underwent a standardised clinical assessment consisting of a medical history, family history, and physical examination. We based symptomatic status on this assessment, which included a collateral history from a family member or close friend. We measured functional status using the Frontotemporal Dementia Rating Scale[27] and assessed behavioural symptoms using the Cambridge Behavioural Inventory Revised version (CBI-R).[28] Patients underwent a neuropsychological battery consisting of tests from the Uniform Data Set:[29] the Logical Memory subtest of the Wechsler Memory Scale-Revised with Immediate and Delayed Recall scores, Digit Span forwards and backwards from the Wechsler Memory Scale-Revised, a Digit Symbol Task, Parts A and B of the Trail Making Test, the short version of the Boston Naming Test, and Category Fluency (animals). We also tested Letter Fluency and did the Wechsler Abbreviated Scale of Intelligence Block Design task, and the Mini-Mental State Examination (MMSE). For each test, apart from the MMSE and CBI-R, we calculated *Z* scores based on language-specific norms. Most at-risk participants (158 [88%] of 180) had not undergone presymptomatic genetic testing and were therefore not aware of their mutation status, and for these participants the clinicians and neuropsychologists who did the assessments were masked to mutation status.

We did volumetric T1-weighted MRI on 3T and 1·5T scanners at sites where 3T scanning was not available. We designed scan protocols at the outset of the study to match across scanners as much as possible. For the volumetric analysis, we did a cortical parcellation using a multiatlas segmentation propagation approach following the brainCOLOR protocol,[30,31] combining regions of interest to calculate grey matter volumes of the entire cortex, separated into the frontal, temporal, parietal, occipital, cingulate, and insula cortices. We also did a subcortical parcellation using the Neuromorphometrics protocol[32,33] for the hippocampus, amygdala, striatum, and thalamus, and a parcellation of the cerebellum using the Diedrichsen cerebellar atlas,[33,34] producing a measure for the entire cerebellum by combining regions of interest. We measured whole-brain volumes using a semi-automated segmentation method.[35] We expressed all measures as a percentage of total intracranial volume (measured with SPM12 with a combination of grey matter, white matter, and CSF segmentations). In view of

previous evidence for asymmetrical atrophy in *GRN* mutation carriers compared with *MAPT* and *C9orf72* carriers,[4,5] we also assessed differences between left and right hemisphere volumes using a laterality index, calculated as the absolute difference between left and right cortical volumes divided by total cortical volume.

Findings from individual case series of individuals with dementia with a known genetic cause suggest that variability of age at symptom onset exists within families. However, authors of a large study of familial Alzheimer's disease[36] suggest that a strong relation exists between individual age at symptoms onset and both parental age at onset and mean age at onset within the family. To our knowledge, no similar studies have been done in frontotemporal dementia. We therefore did an initial analysis on the basis of the symptomatic carriers within our cohort, investigating the relation between their age at symptoms onset and parental age at onset, their age at onset and mean age at onset for other members of the same family, and their age at onset and median age at onset for other members of the same family (excluding the symptomatic individual from mean and median calculations). Parental age at onset did not show a significant correlation with age at symptoms onset of the symptomatic carriers (Pearson correlation coefficient 0·39; p=0·0685), but we found both mean and median ages at onset within the family to be significantly correlated with the symptomatic carriers' age at onset (Pearson correlation coefficient 0·53, p=0·0019 for the mean and 0·50, p=0·0036 for the median). Furthermore, in addition to being correlated with mean age of onset within their families, age at symptoms onset of symptomatic carriers did not significantly differ from mean age at onset within their families (p=0·3216 Wilcoxon signed rank). On the basis of this analysis, we decided to use mean familial age at onset to estimate time to expected symptom onset—ie, someone aged 50 years old at the time of assessment with a mean age at onset of 55 years old in their family would be given an expected time from symptoms onset of –5 years. Data were available for this calculation from one family member in 35 families, from two in 15 families, from three in ten families, from four in four families, from five in five families, from six in two families, and from seven in two families; 12, 16, and 30 family members were available in a further three families.

Department of Neurobiology, Care Sciences and Society, Center for Alzheimer Research, Division of Neurogeriatrics (Prof C Graff MD, A Kinhult Ståhlbom PhD, L Lilius BSc, H Thonberg PhD), **and Division of Clinical Geriatrics** (V Jelic MD), **and Department of Clinical Neuroscience** (C Andersson PhD), **Karolinska Institutet, Huddinge, Sweden; Department of Geriatric**

	Non-carriers (n=102)	Mutation carriers (n=118)
Male	60 (59%)	57 (48%)
Mutated gene		
MAPT	18 (18%)	26 (22%)
GRN	60 (59%)	58 (49%)
C9orf72	24 (24%)	34 (29%)
Clinical status		
Asymptomatic	102 (100%)	78 (66%)
Symptomatic	0	40 (34%)
Right-handed	94 (92%)	106 (90%)
Age (years)	49·2 (36·3–61·7)	53·3 (41·4–62·7)
Education (years)	13 (11–16)	13 (10–16)
Years from expected onset		
–20 or longer	32 (31%)	21 (18%)
–20 up to –10	18 (18%)	21 (18%)
–10 up to 0	23 (23%)	24 (20%)
0 and beyond expected onset	29 (28%)	52 (44%)

Data are n (%) or median (IQR).

Table 1: **Characteristics of study participants**

	–25 years	–20 years	–15 years	–10 years	–5 years	0 years	5 years	10 years
Behavioural								
Cambridge Behavioural Inventory—Revised (/180)								
Non-carriers	1·9	4·6	7·6	10·6	13·2	14·9	15·4	14·2
Carriers	0·2	4·3	9·6	16·2	24·0	33·1	43·5	55·1
Difference	–1·7	–0·3	2·0	5·5	10·8	18·2	28·1	40·9
SE	2·8	2·2	2·7	4·0	4·9	5·3	5·8	8·7
p value	0·5611	0·8867	0·4748	0·1620	0·0269	0·0005	<0·0001	<0·0001
Cognitive								
Mini Mental State Examination (/30)								
Non-carriers	29·5	29·2	28·9	28·6	28·4	28·3	28·3	28·2
Carriers	30·3	29·6	28·8	28·0	27·1	26·1	25·0	23·9
Difference	0·7	0·4	<0·1	–0·6	–1·4	–2·2	–3·2	–4·4
SE	0·3	0·3	0·3	0·4	0·5	0·7	0·9	1·3
p value	0·0221	0·1683	0·9303	0·0922	0·0045	0·0008	0·0006	0·0007
Neuropsychological (*Z* score)								
Logical Memory—Immediate Recall								
Non-carriers	0·4	0·2	0·1	<0·1	–0·1	–0·2	–0·3	–0·4
Carriers	0·4	0·3	0·1	–0·1	–0·4	–0·8	–1·3	–1·9
Difference	<0·1	0·1	<0·1	–0·1	–0·3	–0·6	–1·0	–1·5
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·8948	0·7183	0·8779	0·6136	0·0863	0·0005	<0·0001	<0·0001
Logical Memory—Delayed Recall								
Non-carriers	0·3	0·2	0·1	<0·1	–0·1	–0·2	–0·3	–0·5
Carriers	0·2	0·2	<0·1	–0·2	–0·5	–0·9	–1·3	–1·8
Difference	–0·1	–0·1	–0·1	–0·3	–0·4	–0·7	–1·0	–1·4
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·6463	0·6767	0·4849	0·1696	0·0105	<0·0001	<0·0001	<0·0001
Digit Span forwards								
Non-carriers	0·1	0·1	0·1	<0·1	<0·1	–0·1	–0·1	–0·2
Carriers	0·5	0·3	0·1	–0·2	–0·4	–0·7	–1·0	–1·3
Difference	0·3	0·2	<0·1	–0·2	–0·4	–0·7	–0·9	–1·1
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·1479	0·4366	0·9235	0·2847	0·0253	0·0005	0·0001	0·0003
Digit Span backwards								
Non-carriers	0·1	0·1	<0·1	<0·1	–0·1	–0·1	–0·2	–0·2
Carriers	0·1	<0·1	–0·2	–0·4	–0·6	–0·8	–1·1	–1·4
Difference	–0·1	–0·1	–0·2	–0·3	–0·5	–0·7	–0·9	–1·2
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·8098	0·5866	0·3136	0·0933	0·0079	0·0001	<0·0001	0·0001

(Table 2 continues on next page)

	−25 years	−20 years	−15 years	−10 years	−5 years	0 years	5 years	10 years
(Continued from previous page)								
Digit Symbol Task								
Non-carriers	0·8	0·7	0·5	0·3	0·1	−0·2	−0·4	−0·7
Carriers	0·8	0·6	0·3	<0·1	−0·4	−0·9	−1·4	−1·9
Difference	<0·1	−0·2	−0·2	−0·3	−0·5	−0·7	−0·9	−1·2
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·9036	0·7223	0·3033	0·0549	0·0017	<0·0001	<0·0001	<0·0001
Trail Making Test Part A								
Non-carriers	0·4	0·3	0·2	<0·1	−0·2	−0·4	−0·6	−0·8
Carriers	0·6	0·4	0·1	−0·2	−0·6	−1·0	−1·5	−2·0
Difference	0·2	0·1	−0·1	−0·2	−0·4	−0·6	−0·9	−1·2
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·4662	0·7470	0·7716	0·2832	0·0355	0·0012	0·0002	0·0006
Trail Making Test Part B								
Non-carriers	0·6	0·4	0·3	0·2	<0·1	−0·2	−0·3	−0·5
Carriers	0·9	0·7	0·4	<0·1	−0·5	−1·0	−1·7	−2·5
Difference	0·3	0·2	0·1	−0·2	−0·5	−0·9	−1·4	−1·9
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·1730	0·2639	0·7317	0·3799	0·0072	<0·0001	<0·0001	<0·0001
Letter Fluency								
Non-carriers	0·1	−0·1	−0·2	−0·3	−0·4	−0·5	−0·6	−0·6
Carriers	0·2	0·2	<0·1	−0·3	−0·6	−1·1	−1·7	−2·4
Difference	0·1	0·2	0·2	0·1	−0·2	−0·6	−1·1	−1·8
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·6629	0·3280	0·3592	0·7952	0·2746	0·0015	<0·0001	<0·0001
Category Fluency								
Non-carriers	0·5	0·4	0·2	0·1	−0·1	−0·3	−0·4	−0·6
Carriers	0·6	0·5	0·3	<0·1	−0·4	−0·8	−1·3	−2·0
Difference	0·1	0·1	0·1	−0·1	−0·3	−0·5	−0·9	−1·4
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·6632	0·4945	0·6544	0·7932	0·1226	0·0007	<0·0001	<0·0001
Boston Naming Test								
Non-carriers	<0·1	−0·1	−0·2	−0·3	−0·3	−0·3	−0·3	−0·3
Carriers	0·4	0·2	−0·2	−0·6	−1·0	−1·6	−2·2	−2·9
Difference	0·4	0·3	0·1	−0·3	−0·7	−1·2	−1·9	−2·6
SE	0·3	0·3	0·3	0·3	0·3	0·2	0·3	0·4
p value	0·1763	0·2871	0·7965	0·3202	0·0047	<0·0001	<0·0001	<0·0001
Block Design								
Non-carriers	0·4	0·3	0·2	<0·1	−0·2	−0·3	−0·5	−0·7
Carriers	0·7	0·5	0·2	−0·1	−0·5	−1·0	−1·4	−2·0
Difference	0·3	0·2	<0·1	−0·2	−0·4	−0·6	−0·9	−1·3
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·2220	0·3911	0·8839	0·4029	0·0284	0·0001	<0·0001	<0·0001

Differences calculated from unrounded values.

Table 2: **Behavioural and neuropsychological estimates in mutation carriers and non-carriers, by estimated time from expected symptoms onset**

Medicine (Prof C Graff, M Fallström MSc, V Jelic, A Kinhult Ståhlbom, H Thonberg), **and Department of Psychology** (C Andersson), **Karolinska University Hospital-Huddinge, Stockholm,**

Statistical analysis

We used linear mixed-effects models to examine whether differences existed between non-carriers and mutation carriers in the association between each clinical, behavioural, or structural imaging measure and the time to expected onset of symptoms (we combined all genes because of low numbers in each individual genetic group). This modelling framework allows estimation of fixed and random effects of predictor variables, including the intercept. Fixed effects represent non-random sources of variation, where the predictor variable has the same relation with the outcome in all observations. Random effects estimate the variance in the effect of a predictor between different clusters in the data and this estimation allows for correlation in the outcome between members of the same cluster.[37,38]

For analysis of each measure, a random intercept for family allowed values of the marker to be correlated between family members. The fixed effect predictor variables of interest were mutation carrier status, time to expected onset, and terms for the interaction between mutation carrier status and time to expected onset. We expected a non-linear change in each measure over time, so models also included a quadratic term for time to expected onset and the interaction between this term and mutation carrier status. We included a more complex cubic relation association between the measure and time to expected onset only when significant ($p<0·05$) evidence existed that addition of a cubic term and the interaction between the cubic term and mutation carrier status improved model fit. An example of the mixed effect model is given in the appendix for analysis of whole-brain volume to show the modelling framework that we used for analysis.

We also did exploratory analyses to assess whether differences between non-carriers and *MAPT*, *GRN*, and *C9orf72* mutation carriers existed in the association between values of each measure and time to expected onset of symptoms. Because of the small number of participants in each gene group, we considered only linear changes in markers over time in this analysis.

We did a Wald test for each model to assess whether the mean value of the measure differed between mutation carriers and non-carriers. We predicted average values from the mixed effects model for each group and differences between mutation carriers and non-carriers every 5 years between 25 years before expected onset and 10 years after expected onset. All analyses were adjusted for study site and sex. Model diagnostics for both MMSE and CBI-R suggested non-constant variance, so we used robust standard errors for these analyses.

In addition to the prespecified analysis of markers of disease progression, we did a post-hoc analysis to examine whether differences existed between non-carriers and *MAPT*, *GRN*, and *C9orf72* mutation carriers in the association between laterality of brain volume and time to expected onset of symptoms. Because of strong skew in laterality, we used a log transformation for this analysis, and results are presented as ratios of laterality between mutation carriers and non-carriers for ease of interpretation. We did all analyses with STATA (version 12.1 or later).

Role of the funding source
The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study except for the results of genetic mutation screening in presymptomatic participants. Only JMN and DMC had access to all of the genetic results to avoid risk of disclosure of genetic status to at-risk participants who were unaware of whether they carried a mutation. All authors had final responsibility for the decision to submit for publication.

Results

We analysed data from 220 participants, consisting of 118 mutation carriers and 102 non-carriers (table 1). Of the 118 mutation carriers, 40 were symptomatic (11 with *MAPT*, 13 with *GRN*, and 16 with *C9orf72* mutations) and 78 were asymptomatic (15 with *MAPT*, 45 with *GRN*, and 18 with *C9orf72* mutations). Of the 102 non-carriers, 18 were from families with *MAPT* mutations, 60 were from families with *GRN* mutations, and 24 were from families with *C9orf72* mutations.

Participants came from 76 families (17 with *MAPT*, 32 with *GRN*, and 27 with *C9orf72* mutations), with the mean age at symptom onset across all individuals being 56·9 (SD 8·4) years. Mean age at symptom onset was 49·5 (5·6) years in the *MAPT* families, 57·8 (8·7) years in the *GRN* families, and 60·6 (6·7) years in the *C9orf72* families (appendix). We noted ten different *MAPT* mutations in the 17 families: Pro301Leu, intronic 10+16, Gly272Val, Val363Ile, Arg406Trp, Val337Met, Ser320Phe, Pro301Ser, Leu315Arg, and Gln351Arg (in order of number of participants in study). We found 13 different *GRN* mutations in the 32 families: Ser82fs, Thr272fs, Gln125X, Gln249X, Arg493X, Gln130fs, Cys416fs, Val411fs, Trp386X, Gly35fs, Cys31fs, Cys474fs, and Asp22fs.

In the symptomatic cohort, most participants had a diagnosis of behavioural variant frontotemporal dementia (meeting the Rascovsky diagnostic criteria),[39] except for six participants with *GRN* mutations who had diagnoses of the non-fluent variant of primary progressive aphasia (Gorno-Tempini diagnostic criteria)[40] and four participants with *C9orf72* mutations (one with the non-fluent variant of primary progressive aphasia, two with frontotemporal dementia with motor neuron disease, and one with a dementia syndrome not otherwise specified). Functionally, one participant (with a *MAPT* mutation) in the symptomatic cohort was very mildly affected (according to the Frontotemporal Dementia Rating Scale), three (one *GRN* and two *C9orf72*) were mildly affected, 16 (four *MAPT*, five *GRN*, and seven *C9orf72*) were moderately affected, 13 (four *MAPT*, four *GRN*, and five *C9orf72*) were severely affected, and seven (two *MAPT*, three *GRN*, and two *C9orf72*) were very severely affected.

MMSE, CBI-R, and all neuropsychology measures showed significant mean differences between mutation

	–25 years	–20 years	–15 years	–10 years	–5 years	0 years	5 years	10 years
Whole-brain volume (% of TIV)								
Whole brain								
Non-carriers	86·1%	84·8%	83·7%	82·6%	81·6%	80·7%	79·8%	79·0%
Carriers	87·0%	85·7%	84·0%	82·1%	79·8%	77·2%	74·3%	71·1%
Difference	0·9%	0·8%	0·3%	–0·5%	–1·8%	–3·5%	–5·5%	–8·0%
SE	0·9	0·8	0·8	0·8	0·7	0·7	0·9	1·2
p value	0·3184	0·3198	0·6738	0·5004	0·0157	<0·0001	<0·0001	<0·0001
Cortical volume (% of TIV)								
Frontal lobe								
Non-carriers	12·9%	12·7%	12·4%	12·2%	12·0%	11·8%	11·7%	11·5%
Carriers	13·1%	12·8%	12·4%	12·0%	11·5%	11·1%	10·5%	10·0%
Difference	0·2%	0·1%	<0·1%	–0·2%	–0·5%	–0·8%	–1·1%	–1·5%
SE	0·2	0·2	0·2	0·2	0·2	0·2	0·2	0·3
p value	0·3208	0·5261	0·8689	0·1766	0·0023	<0·0001	<0·0001	<0·0001
Temporal lobe								
Non-carriers	8·8%	8·6%	8·5%	8·3%	8·2%	8·1%	8·0%	7·9%
Carriers	8·7%	8·6%	8·4%	8·1%	7·9%	7·6%	7·3%	6·9%
Difference	<0·1%	<0·1%	–0·1%	–0·2%	–0·3%	–0·5%	–0·7%	1·0%
SE	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·2
p value	0·8944	0·7049	0·3287	0·0483	0·0005	<0·0001	<0·0001	<0·0001
Parietal lobe								
Non-carriers	7·1%	7·0%	6·8%	6·7%	6·6%	6·4%	6·4%	6·3%
Carriers	7·0%	6·9%	6·7%	6·6%	6·4%	6·2%	5·9%	5·6%
Difference	–0·1%	–0·1%	–0·1%	–0·1%	–0·2%	–0·3%	–0·5%	–0·6%
SE	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·2
p value	0·2818	0·4546	0·4800	0·2820	0·0510	0·0010	<0·0001	<0·0001
Occipital lobe								
Non-carriers	5·6%	5·6%	5·5%	5·4%	5·4%	5·3%	5·2%	5·1%
Carriers	5·6%	5·5%	5·5%	5·4%	5·3%	5·2%	5·0%	4·9%
Difference	–0·1%	–0·1%	<0·1%	–0·1%	–0·1%	–0·1%	–0·2%	–0·3%
SE	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
p value	0·4022	0·5072	0·5377	0·4311	0·2181	0·0554	0·0166	0·0175
Insula								
Non-carriers	0·86%	0·85%	0·85%	0·84%	0·83%	0·82%	0·80%	0·79%
Carriers	0·85%	0·84%	0·82%	0·80%	0·77%	0·74%	0·71%	0·67%
Difference	–0·01%	–0·02%	–0·03%	–0·04%	–0·05%	–0·07%	–0·10%	–0·12%
SE	0·02	0·02	0·02	0·02	0·02	0·01	0·02	0·02
p value	0·5379	0·2992	0·1028	0·0131	0·0002	<0·0001	<0·0001	<0·0001
Cingulate								
Non-carriers	1·95%	1·91%	1·89%	1·86%	1·84%	1·82%	1·81%	1·79%
Carriers	1·98%	1·95%	1·91%	1·86%	1·80%	1·74%	1·67%	1·59%
Difference	0·04%	0·03%	0·02%	<0·01%	–0·04%	–0·08%	–0·14%	–0·20%
SE	0·04	0·03	0·03	0·03	0·03	0·03	0·03	0·05
p value	0·3386	0·3246	0·5478	0·8934	0·1935	0·0036	<0·0001	<0·0001
Subcortical volume (% of TIV)								
Hippocampus								
Non-carriers	0·70%	0·69%	0·68%	0·68%	0·67%	0·66%	0·65%	0·64%
Carriers	0·69%	0·69%	0·68%	0·66%	0·64%	0·62%	0·59%	0·55%
Difference	–0·01%	–0·01%	–0·01%	–0·01%	–0·02%	–0·04%	–0·06%	–0·09%
SE	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·02
p value	0·4421	0·6667	0·6464	0·3408	0·0441	0·0003	<0·0001	<0·0001

(Table 3 continues on next page)

	–25 years	–20 years	–15 years	–10 years	–5 years	0 years	5 years	10 years
(Continued from previous page)								
Amygdala								
Non-carriers	0·14%	0·14%	0·14%	0·14%	0·14%	0·14%	0·14%	0·14%
Carriers	0·14%	0·14%	0·14%	0·14%	0·13%	0·13%	0·12%	0·12%
Difference	<0·01%	<0·01%	<0·01%	<0·01%	–0·01%	–0·01%	–0·01%	–0·02%
SE	0·003	0·003	0·003	0·003	0·003	0·003	0·003	0·005
p value	0·7016	0·7182	0·5451	0·2397	0·0302	0·0005	<0·0001	<0·0001
Striatum								
Non-carriers	1·28%	1·26%	1·25%	1·24%	1·23%	1·23%	1·22%	1·22%
Carriers	1·29%	1·26%	1·23%	1·21%	1·18%	1·16%	1·13%	1·11%
Difference	0·01%	<0·01%	–0·01%	–0·03%	–0·05%	–0·07%	–0·09%	–0·11%
SE	0·03	0·02	0·02	0·02	0·02	0·02	0·02	0·03
p value	0·6840	0·9889	0·5552	0·1928	0·0255	0·0010	0·0002	0·0008
Thalamus								
Non-carriers	0·95%	0·93%	0·92%	0·90%	0·89%	0·88%	0·87%	0·85%
Carriers	0·94%	0·92%	0·91%	0·89%	0·86%	0·83%	0·80%	0·76%
Difference	–0·02%	–0·01%	–0·01%	–0·02%	–0·03%	–0·04%	–0·07%	–0·09%
SE	0·02	0·02	0·02	0·01	0·01	0·01	0·02	0·02
p value	0·3630	0·4977	0·4688	0·2472	0·0385	0·0007	<0·0001	<0·0001
Cerebellar volume (% of TIV)								
Cerebellar								
Non-carriers	7·6%	7·5%	7·4%	7·3%	7·2%	7·1%	7·1%	7·0%
Carriers	7·6%	7·6%	7·5%	7·4%	7·3%	7·1%	6·9%	6·6%
Difference	0·1%	0·1%	0·2%	0·1%	0·1%	<0·1%	–0·2%	–0·4%
SE	0·1	0·1	0·1	0·1	0·1	0·1	0·1	0·1
p value	0·5840	0·1865	0·1070	0·1478	0·4216	0·6560	0·0604	0·0071

Differences calculated from unrounded values. TIV=total intracranial volume.

***Table 3*: Imaging estimates in mutation carriers and non-carriers, by estimated time from expected symptoms onset**

Sweden; Istituto di Ricovero e Cura a Carattere Scientifico Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy (Prof G B Frisoni MD, M Pievani PhD, M Bocchetta, L Benussi PhD, R Ghidoni PhD); **Memory Clinic and LANVIE—Laboratory of Neuroimaging of Aging, University Hospitals and University of Geneva, Geneva, Switzerland** (Prof G B Frisoni); **Department of Clinical Neurological Sciences, University of Western Ontario, London, ON, Canada** (E Finger MD); **and Department of Neuroscience, Psychology, Drug Research and Child Health** (Prof S Sorbi MD, B Nacmias PhD, G Lombardi MD), **and Department of Clinical Pathophysiology, Nuclear Medicine Division** (C Polito PhD), **University of Florence, Florence, Italy**

carriers as a whole group and non-carriers (p≤0·0028 for all markers). MMSE, CBI-R, and all neuropsychology measures except the Logical Memory Immediate Recall and verbal fluency tasks showed significant mean differences between mutation carriers as a whole group and non-carriers 5 years before expected onset (table 2 and appendix). We noted no significant differences at timepoints earlier than 5 years before expected onset. The earliest point at which the Logical Memory Immediate Recall and verbal fluency tasks showed differences between mutation carriers and non-carriers was at the time of expected onset. In the exploratory analysis of individual genetic groups, the behavioural and neuropsychological tests that showed differences between mutation carriers and non-carriers at the earliest times before expected onset were different in each genetic group: the Boston Naming Test and the CBI-R for the *MAPT* group, the Digit Span backwards for the *GRN* group, and the CBI-R in the *C9orf72* group (appendix).

We did volumetric T1-weighted MRI in 212 participants (eight were unable to have a scan because of either contraindications to MRI scanning or claustrophobia). A further ten scans did not pass an initial quality control process, usually owing to excessive motion during the scan. We therefore used 202 scans for analysis (175 from 3T scanners [55 Siemens, 99 Philips, and 21 General Electric scanners] and 27 from 1·5T scanners [19 Siemens and 8 General Electric scanners]). 93 scans were from non-carriers and 109 from mutation carriers (24 *MAPT,* 52 *GRN,* and 33 *C9orf72*). Whole-brain volume showed a significant difference between mutation carriers as a whole group and non-carriers (p<0·0001), with strong evidence for a difference in all cortical and subcortical volumes (p≤0·0030), except for the occipital lobe, which was not significant (p=0·0598). The cerebellum had a less significant difference than the cortical and subcortical volumes (p=0·0211). We noted differences in group means between mutation carriers and non-carriers at the earliest timepoint for the insula (10 years before expected symptom onset) followed by the temporal lobe (also 10 years before expected symptom onset, but with a less significant difference; table 3 and figure). We noted differences in the frontal lobe, all subcortical volumes, and whole-brain volume between carriers and non-carriers at 5 years before expected onset, whereas we noted differences in the parietal lobe and cingulate only just before expected time of onset (table 3, figure, and appendix). Although we noted only weak evidence for a difference between mutation carriers and non-carriers, the results suggest that significant differences might exist in the occipital lobe at 5 years after symptoms onset and in the cerebellum at 10 years after symptoms onset.

When we analysed the individual genetic groups separately, we noted a different ordering of cortical and subcortical involvement in each group (appendix): in the *MAPT* group, we noted differences between mutation carriers and non-carriers in the hippocampus and amygdala at 15 years before expected onset, followed by the temporal lobe at 10 years before expected onset, and the insula at 5 years before expected onset; in the *GRN* group, we noted differences between carriers and non-carriers in the insula at 15 years before expected onset, then in the temporal and parietal lobes at 10 years before expected onset, with the earliest subcortical area affected being the striatum at 5 years before expected onset; and in the *C9orf72* group, subcortical areas including the thalamus, the insula, and posterior cortical areas differed between carriers and controls at 25 years before expected onset, followed by the frontal and temporal lobes at 20 years before expected onset. We noted significant differences in the cerebellum presymptomatically in the *C9orf72* group at 10 years before expected onset. Examination of the laterality index showed evidence for asymmetry between left and right cortical volumes in the *GRN* mutation carriers (p=0·0001 *vs* non-carriers), but not in the *MAPT* carriers (p=0·3283 *vs* non-carriers) or *C9orf72* carriers (p=0·2018 *vs* non-carriers). *GRN* mutation carriers showed significantly greater asymmetry than non-carriers at 5 years before expected onset (appendix).

Discussion

We have shown that imaging changes can be identified at least 10 years before expected onset of symptoms in genetic frontotemporal dementia. Structural neuroimaging identifies a sequence of change in atrophy through cortical and subcortical regions, with the insular and temporal cortices affected initially (around 10 years before expected symptoms onset), followed by the frontal cortex and subcortical areas (around 5 years before expected onset), parietal and cingulate cortices (around time of expected onset), and, lastly, the occipital cortex (5 years after expected onset) and cerebellum (10 years after expected onset). We noted that neuropsychological measures were first different between carriers and non-carriers later than initial imaging measures, up to 5 years before expected symptoms onset. These findings suggest that the disease process significantly precedes onset of symptoms in genetic frontotemporal dementia. Whereas previous studies have shown inconsistent findings (panel), the value of investigation of a large cohort of presymptomatic participants is confirmed in this study, consistent with similar approaches previously done in patients with familial Alzheimer's disease[8] and patients with Huntington's disease.[7]

The findings from this study are consistent with our understanding of the earliest structural changes in frontotemporal dementia. The insula is thought to act as a crucial hub in many key networks that become affected (particularly the so-called salience network connecting the insula, frontal lobe, and anterior cingulate, and frontoparietal networks).[25,41,42] Here, we noted that the insula was the first cortical area to show evidence of atrophy in the mutation group as a whole, and was one of the earliest areas affected in the analyses of each individual genetic group, suggesting that it might be an early focus of pathology followed by connectivity-based spread of disease.

Our primary analysis focused on genetic frontotemporal dementia as a single group. The rationale for this decision lies in the shared clinical features and overlapping disease mechanisms seen in genetic frontotemporal dementia. However, differences have been shown between genetic subgroups in previous neuroimaging studies,[43,44] and signatures of network disintegration with particular genetic proteinopathies are predicted on both empirical and theoretical grounds.[45] Our exploratory analyses are consistent with and extend this previous work. In the *MAPT* group, temporal lobe and medial temporal structures (the hippocampus and amygdala) were affected initially, consistent with previous findings suggesting that the disease is a temporal-predominant disorder.[18,43,46] However, this study shows that significant changes can be seen in these areas much earlier than previously suggested. In the *GRN* group, the insula was the first area affected (around 15 years before expected onset), followed by the temporal and parietal lobes. Consistent with previous neuroimaging studies of symptomatic carriers showing early temporal and parietal involvement in patients with *GRN* mutations,[11,43,46] findings from this study identify the insula as the key region affected significantly earlier than other areas. Distinct from the other groups, the earliest subcortical involvement in the *GRN* group was in the striatum (around 5 years before expected onset), an area known to be involved in symptomatic *GRN* mutation cases, but not previously shown presymptomatically.[47] In the *C9orf72* group, the thalamus and more posterior cortical areas were affected early. No previous presymptomatic studies of this group have been done, but previous imaging analyses of symptomatic carriers suggest that the thalamus is a key area affected in people with *C9orf72* expansions and that posterior areas are more involved than in the other two genetic groups.[43,44] Similarly, the cerebellum has been identified as an area affected in symptomatic *C9orf72* expansion carriers, and here we show evidence for presymptomatic involvement. The exploratory analysis suggested very early detectable structural imaging changes, particularly in the *C9orf72* group, more than 20 years before expected symptoms onset. The timing of

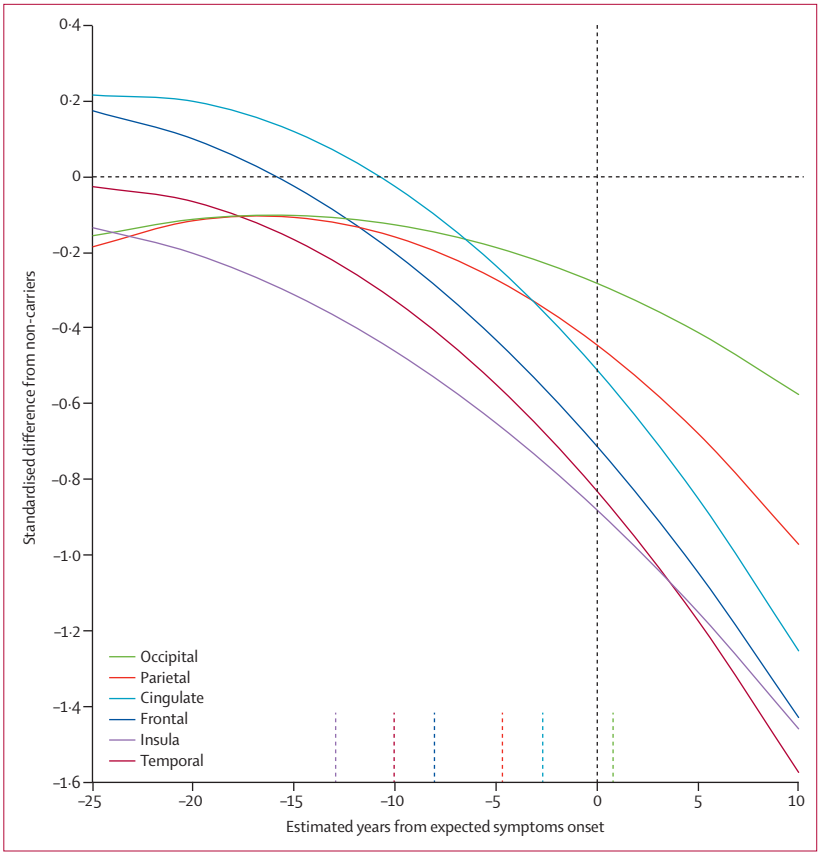

Figure: **Standardised difference between all mutation carriers and non-carriers in cortical grey matter volumetric imaging measures versus estimated years from expected symptoms onset**
Individual datapoints not plotted to prevent disclosure of genetic status. The time at which the upper 95% CI for each curve crosses zero on the y-axis (ie, the point at which a significant difference exists between mutation carriers and non-carriers) is shown on the x-axis. Individual curves with 95% CIs are shown in the appendix. Subcortical and cerebellar volumes are also shown in the appendix.

Correspondence to:
Prof Martin Rossor, Dementia Research Centre, Department of Neurodegenerative Disease, University College London Institute of Neurology, University College London, London WC1N 3BG, UK
m.rossor@ucl.ac.uk

See **Online** for appendix

For the **Genetic Frontotemporal dementia Initiative** see www.genfi.org.uk

Panel: Research in context

Systematic review

We searched PubMed for articles on presymptomatic studies in genetic frontotemporal dementia up to Nov 16, 2014, using the following terms: "frontotemporal dementia AND genetics" and "frontotemporal dementia AND presymptomatic". We identified one review article of presymptomatic studies in genetic frontotemporal dementia,[5] and 18 original research studies that had investigated neuropsychology or neuroimaging, or both, in presymptomatic genetic frontotemporal dementia (appendix)[9–21] A few case studies[9,11,16,17] and two other studies[15,22] have shown evidence of presymptomatic abnormalities on neuropsychometry in asymptomatic mutation carriers, usually with tests of executive dysfunction. However, findings from some other studies have not shown any abnormalities before onset.[13,19,21,23–25] In two single case studies[9,11] and two small case series[12,13] of presymptomatic *GRN* mutation carriers, focal brain atrophy has been shown a few years before symptoms onset using volumetric T1 MRI, with the prefrontal cortex being predominantly involved, often in an asymmetric pattern. *MAPT* carriers have been studied less than *GRN* carriers, with a single case study[17] and a small case series[18] showing presymptomatic atrophy, with hippocampal involvement predominating. We identified no presymptomatic studies of *C9orf72* mutation carriers. Some studies have focused on other types of MRI in *GRN* and *MAPT* carriers, particularly diffusion tensor imaging and resting-state functional MRI;[19–26] however, Borroni and colleagues,[19,21] Whitwell and colleagues,[20] and Dopper and colleagues[22] also did voxel-based morphometry analyses using volumetric T1 imaging in their studies and did not find any differences between asymptomatic carriers and controls.

Interpretation

This work is the first multicentre study of presymptomatic genetic frontotemporal dementia and identifies structural imaging changes around 10 years before expected onset, and cognitive impairment around 5 years before expected onset, when the genetic group is investigated as a whole. Exploratory analyses suggest that different cortical and subcortical areas are affected earliest in each of the *MAPT*, *GRN*, and *C9orf72* groups, and that structural imaging changes can be seen 15 years or more before symptoms onset. Our results provide an insight into the early neuroanatomical changes in genetic frontotemporal dementia and suggest the potential for use of structural imaging measures as biomarkers in future therapeutic trials.

presymptomatic involvement before expected symptoms onset might, to some extent, result from limitations of the simple linear association used in modelling, but this intriguing finding needs further investigation and could be consistent with the very slow progressive change in symptoms seen in some patients with *C9orf72*-related frontotemporal dementia.[48–50] Another possibility is that some of the very early differences between mutation carriers and non-carriers in the *C9orf72* group represent differences in brain volume that are, in fact, developmental and longstanding, with superimposed atrophy only late in the disease process.

A key strength of this study is its ability to show robust presymptomatic differences in clinical and imaging biomarkers in genetic frontotemporal dementia. However, we analysed only cross-sectional differences between carriers and controls at different times from expected symptoms onset. Whether the apparent progression of atrophy through a sequence of cortical and subcortical regions is followed within individuals remains to be shown in a longitudinal study. A further limitation of the study is the method used for estimation of age at onset in presymptomatic mutation carriers. Despite our initial analysis showing a significant correlation between actual age at onset in symptomatic carriers and mean familial age at onset, this measure is imperfect, with variability in age at onset within a family in all frontotemporal dementia mutations. This variability is greater for *C9orf72* and *GRN* mutations than for *MAPT* mutations, which could lead to greater error in estimated time to onset in these subtypes than in the *MAPT* subtype (and could therefore suggest that changes can be seen earlier than actually occur). Another limitation of the study is its ability to detect subtle neuropsychiatric or neuropsychological abnormalities. The behavioural and cognitive battery used in the study includes a series of standard validated tests, but these tests might not have sufficient sensitivity for diagnosis of subtle cognitive or neuropsychiatric dysfunction identified with experimental tests.

In further studies, imaging, genetic, biochemical, and cognitive measures might be able to be combined to identify changes even earlier than noted here. Findings from initial studies[19–25] suggest that presymptomatic differences between carriers and non-carriers of mutations associated with frontotemporal dementia might be seen with other imaging methods, such as diffusion tensor imaging and resting-state functional MRI. Findings from presymptomatic studies of Alzheimer's disease[8] also suggest earlier changes in ^{11}C Pittsburgh compound B PET and CSF measures than diffusion tensor imaging and resting-state functional MRI. Although no fluid biomarkers have been identified for frontotemporal dementia, tau PET scanning is now available[51] and will be important to examine`` in this cohort as the GENFI study progresses. Our findings suggest that some readily measurable markers can show rates of decline before symptom onset in frontotemporal dementia; if confirmed in the longitudinal stages of the GENFI study, these measures could be suitable for use in clinical trials and, we hope, contribute to development of preventive strategies.

Contributors

JDR drafted the initial version of the report and the figures. JMN did the statistical analysis. RvM, SM, ER, HT, LB, GB, and BN did genetic analyses. All authors recruited patients, collected data, and contributed to reviewing and editing of the report.

Declaration of interests

JDR is funded by a National Institute for Health Research Rare Disease Translational Research Collaboration Fellowship. SAR is supported by a Vici grant from the Netherlands Organization for Scientific Research. MM is supported by the Department of Medicine, Sunnybrook Health Sciences Centre, and the Sunnybrook Foundation. He is funded by the Canadian Institutes of Health Research and the Ontario Research Fund. He reports consultancy fees from Novartis, Teva, Union Chimique Belge Canada, General Electric Healthcare, and Bioscape Medical Imaging CRO all outside the submitted work. He also holds a US patent of a pharmacogenetic test for Parkinson's disease. MF receives support from the Saul A Silverman Family Foundation as a Canada International Scientific Exchange Program and Morris Kerzner Memorial Fund, and is listed on a provisional patent related to methods and kits for differential

diagnosis of Alzheimer's disease versus frontotemporal dementia with blood biomarkers. ER is funded by the Weston Brain Institute and Ontario Research Fund. RL Jr is funded by Réseau de médecine génétique appliquée, Fonds de recherche du Québec—Santé (FRQS). FT, RG, and LB are funded by the Italian Ministry of Health. DG, ES, AA, and GF are supported by the Fondazione Monzino and Italian Ministry of Health, Ricerca Corrente. JBR is supported by a Wellcome Trust Senior Research Fellowship (088324). JBR and IC-G are supported by the National Institute for Health Cambridge Biomedical Research Centre and Biomedical Research Unit in Dementia. BN is funded by Cassa di Risparmio di Pistoia e Pescia (CRPT 2013/0347). SS is funded by Cassa di Risparmio di Firenze (CRF 2013/0199) and the Ministry of Health RF-2010-2319722. JDW is funded by a Wellcome Trust Senior Clinical Fellowship (091673/Z/10/Z). NCF and MNR are National Institute for Health Research Senior Investigators. NCF also reports consultancy fees (all paid to University College London) from Janssen/Pfizer, General Electric Healthcare, IXICO, Johnson & Johnson, Genzyme (Sonofi company), Eisai, Janssen Alzheimer's Immunotherapy Research and Development, Lilly Research Laboratories (AVID), Eli Lilly, and Novartis Pharma AG all outside the submitted work. He also has a patent QA Box issued. The Dementia Research Centre at University College London is an Alzheimer's Research UK coordinating centre and has received equipment funded by Alzheimer's Research UK and Brain Research Trust. MNR also reports fees (paid to University College London) for serving on a Data Monitoring Committee for Servier outside the submitted work. SEB reports personal fees from Pfizer, GlaxoSmithKline, Novartis, Roche, Eli Lilly, Pfizer, Eisai, Boehringer Ingelheim, General Electric Healthcare, and Novartis, and grants from Pfizer, Elan/Transition Therapeutics Ireland Ltd, Roche, Eli Lilly, General Electric Healthcare, and Lundbeck all outside the submitted work. SO is funded by the Engineering and Physical Sciences Research Council (EP/H046410/1, EP/J020990/1, EP/K005278), the Medical Research Council (MR/J01107X/1), the EU-FP7 project VPH-DARE@IT (FP7-ICT-2011-9-601055), and the National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR BRC UCLH/UCL High Impact Initiative BW.mn.BRC10269). All other authors declare no competing interests.

Acknowledgments

This work was funded by the UK Medical Research Council, the Italian Ministry of Health, and the Canadian Institutes of Health Research as part of a Centres of Excellence in Neurodegeneration grant. JDR, JDW, NCF, and MNR acknowledge the support of the National Institute for Health Research Queen Square Dementia Biomedical Research Unit, Leonard Wolfson Experimental Neurology Centre, the Brain Research Trust, and the University College London Hospitals NHS Trust Biomedical Research Centre. CG acknowledges the support of Swedish Brain Power. ME and SO acknowledge the support of VPH-DARE@IT. We thank Daniel Marcus and Rick Herrick of Radiologics for their support in setting up of the GENFI XNAT database. We would like to thank the participants and their families for taking part in the GENFI study.

References

1 Seelaar H, Rohrer JD, Pijnenburg YA, Fox NC, van Swieten JC. Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. *J Neurol Neurosurg Psychiatry* 2011; **82:** 476–86.

2 Onyike CU, Diehl-Schmid J. The epidemiology of frontotemporal dementia. *Int Rev Psychiatry* 2013; **25:** 130–37.

3 Rohrer JD, Guerreiro R, Vandrovcova J, et al. The heritability and genetics of frontotemporal lobar degeneration. *Neurology* 2009; **73:** 1451–56.

4 Warren JD, Rohrer JD, Rossor MN. Clinical review. Frontotemporal dementia. *BMJ* 2013; **347:** f4827.

5 Rohrer JD, Warren JD, Fox NC, Rossor MN. Presymptomatic studies in genetic frontotemporal dementia. *Rev Neurol (Paris)* 2013; **169:** 820–24.

6 Scahill RI, Schott JM, Stevens JM, Rossor MN, Fox NC. Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. *Proc Natl Acad Sci USA* 2002; **99:** 4703–07.

7 Tabrizi SJ, Langbehn DR, Leavitt BR, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurol* 2009; **8:** 791–801.

8 Bateman RJ, Xiong C, Benzinger TL, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med* 2012; **367:** 795–804.

9 Janssen JC, Schott JM, Cipolotti L, et al. Mapping the onset and progression of atrophy in familial frontotemporal lobar degeneration. *J Neurol Neurosurg Psychiatry* 2005; **76:** 162–68.

10 Rohrer JD, Beck J, Plagnol V, et al. Exome sequencing reveals a novel partial deletion in the progranulin gene causing primary progressive aphasia. *J Neurol Neurosurg Psychiatry* 2013; **84:** 1411–12.

11 Rohrer JD, Warren JD, Barnes J, et al. Mapping the progression of progranulin-associated frontotemporal lobar degeneration. *Nat Clin Pract Neurol* 2008; **4:** 455–60.

12 Cruchaga C, Fernández-Seara MA, Seijo-Martínez M, et al. Cortical atrophy and language network reorganization associated with a novel progranulin mutation. *Cereb Cortex* 2009; **19:** 1751–60.

13 Pievani M, Paternicò D, Benussi L, et al. Pattern of structural and functional brain abnormalities in asymptomatic granulin mutation carriers. *Alzheimers Dement* 2014; **10** (5 suppl)**:** S354–63.e1.

14 Jacova C, Hsiung GY, Tawankanjanachot I, et al. Anterior brain glucose hypometabolism predates dementia in progranulin mutation carriers. *Neurology* 2013; **81:** 1322–31.

15 Hallam BJ, Jacova C, Hsiung GY, et al. Early neuropsychological characteristics of progranulin mutation carriers. *J Int Neuropsychol Soc* 2014; **20:** 694–703.

16 Ferman TJ, McRae CA, Arvanitakis Z, Tsuboi Y, Vo A, Wszolek ZK. Early and pre-symptomatic neuropsychological dysfunction in the PPND family with the N279K tau mutation. *Parkinsonism Relat Disord* 2003; **9:** 265–70.

17 Spina S, Farlow MR, Unverzagt FW, et al. The tauopathy associated with mutation +3 in intron 10 of Tau: characterization of the MSTD family. *Brain* 2008; **131** (Pt 1)**:** 72–89.

18 Miyoshi M, Shinotoh H, Wszolek ZK, et al. In vivo detection of neuropathologic changes in presymptomatic *MAPT* mutation carriers: a PET and MRI study. *Parkinsonism Relat Disord* 2010; **16:** 404–08.

19 Borroni B, Alberici A, Premi E, et al. Brain magnetic resonance imaging structural changes in a pedigree of asymptomatic progranulin mutation carriers. *Rejuvenation Res* 2008; **11:** 585–95.

20 Whitwell JL, Josephs KA, Avula R, et al. Altered functional connectivity in asymptomatic MAPT subjects: a comparison to bvFTD. *Neurology* 2011; **77:** 866–74.

21 Borroni B, Alberici A, Cercignani M, et al. Granulin mutation drives brain damage and reorganization from preclinical to symptomatic FTLD. *Neurobiol Aging* 2012; **33:** 2506–20.

22 Dopper EG, Rombouts SA, Jiskoot LC, et al. Structural and functional brain connectivity in presymptomatic familial frontotemporal dementia. *Neurology* 2013; **80:** 814–23.

23 Premi E, Gazzina S, Bozzali M, et al. Cognitive reserve in granulin-related frontotemporal dementia: from preclinical to clinical stages. *PLoS One* 2013; **8:** e74762.

24 Premi E, Cauda F, Gasparotti R, et al. Multimodal FMRI resting-state functional connectivity in granulin mutations: the case of fronto-parietal dementia. *PLoS One* 2014; **9:** e106500.

25 Premi E, Formenti A, Gazzina S, et al. Effect of TMEM106B polymorphism on functional network connectivity in asymptomatic *GRN* mutation carriers. *JAMA Neurol* 2014; **71:** 216–21.

26 Kantarci K, Boeve BF, Wszolek ZK, et al. MRS in presymptomatic *MAPT* mutation carriers: a potential biomarker for tau-mediated pathology. *Neurology* 2010; **75:** 771–78.

27 Mioshi E, Hsieh S, Savage S, Hornberger M, Hodges JR. Clinical staging and disease progression in frontotemporal dementia. *Neurology* 2010; **74:** 1591–97.

28 Wear HJ, Wedderburn CJ, Mioshi E, et al. The Cambridge Behavioural Inventory revised. *Dementia Neuropsychologia* 2008; **2:** 102–07.

29 Morris JC, Weintraub S, Chui HC, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Dis Assoc Disord* 2006; **20:** 210–16.

30 Jorge Cardoso M, Leung K, Modat M, et al. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. *Med Image Anal* 2013; **17:** 671–84.

31 brainCOLOR. Cortical parcellation (anatomical labeling) protocol. http://www.braincolor.org/protocols/docs/BrainCOLOR_cortical_parcellation_protocol.pdf (accessed Nov 14, 2013).

32 Neuromorphometrics, Inc. Segmentation Protocols. Jan 16, 2007. http://www.neuromorphometrics.org:8080/seg (accessed Nov 14, 2013).

33 Cardoso MJ, Wolz R, Modat M, Fox NC, Rueckert D, Ourselin S. Geodesic information flows. *Med Image Comput Comput Assist Interv* 2012; **15** (Pt 2)**:** 262–70.

34 Diedrichsen J, Balsters JH, Flavell J, Cussans E, Ramnani N. A probabilistic MR atlas of the human cerebellum. *Neuroimage* 2009; **46:** 39–46.

35 Freeborough PA, Fox NC, Kitney RI. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Comput Methods Programs Biomed* 1997; **53:** 15–25.

36 Ryman DC, Acosta-Baena N, Aisen PS, et al. Symptom onset in autosomal dominant Alzheimer disease: a systematic review and meta-analysis. *Neurology* 2014; **83:** 253–60.

37 Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38:** 963–74.

38 Sullivan LM, Dukes KA, Losina E. An introduction to hierarchical linear modelling. *Stat Med* 1999; **18:** 855–88.

39 Rascovsky K, Hodges JR, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 2011; **134** (Pt 9)**:** 2456–77.

40 Gorno-Tempini ML, Hillis AE, Weintraub S, et al. Classification of primary progressive aphasia and its variants. *Neurology* 2011; **76:** 1006–14.

41 Seeley WW, Menon V, Schatzberg AF, et al. Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci* 2007; **27:** 2349–56.

42 Zhou J, Greicius MD, Gennatas ED, et al. Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. *Brain* 2010; **133** (Pt 5)**:** 1352–67.

43 Whitwell JL, Weigand SD, Boeve BF, et al. Neuroimaging signatures of frontotemporal dementia genetics: C9ORF72, tau, progranulin and sporadics. *Brain* 2012; **135** (Pt 3)**:** 794–806.

44 Mahoney CJ, Beck J, Rohrer JD, et al. Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. *Brain* 2012; **135** (Pt 3)**:** 736–50.

45 Warren JD, Rohrer JD, Schott JM, Fox NC, Hardy J, Rossor MN. Molecular nexopathies: a new paradigm of neurodegenerative disease. *Trends Neurosci* 2013; **36:** 561–69.

46 Rohrer JD, Ridgway GR, Modat M, et al. Distinct profiles of brain atrophy in frontotemporal lobar degeneration caused by progranulin and tau mutations. *Neuroimage* 2010; **53:** 1070–76.

47 Premi E, Garibotto V, Gazzina S, et al. Subcortical and deep cortical atrophy in frontotemporal dementia due to granulin mutations. *Dement Geriatr Cogn Dis Extra* 2014; **4:** 95–102.

48 Khan BK, Yokoyama JS, Takada LT, et al. Atypical, slowly progressive behavioural variant frontotemporal dementia associated with C9ORF72 hexanucleotide expansion. *J Neurol Neurosurg Psychiatry* 2012; **83:** 358–64.

49 Suhonen NM, Kaivorinne AL, Moilanen V, et al. Slowly progressive frontotemporal lobar degeneration caused by the C9ORF72 repeat expansion: a 20-year follow-up study. *Neurocase* 2015; **21:** 85–89.

50 Gómez-Tortosa E, Serrano S, de Toledo M, Pérez-Pérez J, Sainz MJ. Familial benign frontotemporal deterioration with C9ORF72 hexanucleotide expansion. *Alzheimers Dement* 2014; **10** (5 suppl)**:** S284–89.

51 Villemagne VL, Okamura N. In vivo tau imaging: obstacles and progress. *Alzheimers Dement* 2014; **10** (3 suppl)**:** S254–64.

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 MARCH 31, 2016 VOL. 374 NO. 13

Randomized Trial of Longer-Term Therapy for Symptoms Attributed to Lyme Disease

Anneleen Berende, M.D., Hadewych J.M. ter Hofstede, M.D., Ph.D., Fidel J. Vos, M.D., Ph.D., Henriët van Middendorp, Ph.D., Michiel L. Vogelaar, M.Sc., Mirjam Tromp, Ph.D., Frank H. van den Hoogen, M.D., Ph.D., A. Rogier T. Donders, Ph.D., Andrea W.M. Evers, Ph.D., and Bart Jan Kullberg, M.D., Ph.D.

ABSTRACT

BACKGROUND

The treatment of persistent symptoms attributed to Lyme disease remains controversial. We assessed whether longer-term antibiotic treatment of persistent symptoms attributed to Lyme disease leads to better outcomes than does shorter-term treatment.

METHODS

In a randomized, double-blind, placebo-controlled trial conducted in Europe, we assigned patients with persistent symptoms attributed to Lyme disease — either related temporally to proven Lyme disease or accompanied by a positive IgG or IgM immunoblot assay for *Borrelia burgdorferi* — to receive a 12-week oral course of doxycycline, clarithromycin plus hydroxychloroquine, or placebo. All study groups received open-label intravenous ceftriaxone for 2 weeks before initiating the randomized regimen. The primary outcome measure was health-related quality of life, as assessed by the physical-component summary score of the RAND-36 Health Status Inventory (RAND SF-36) (range, 15 to 61, with higher scores indicating better quality of life), at the end of the treatment period at week 14, after the 2-week course of ceftriaxone and the 12-week course of the randomized study drug or placebo had been completed.

RESULTS

Of the 281 patients who underwent randomization, 280 were included in the modified intention-to-treat analysis (86 patients in the doxycycline group, 96 in the clarithromycin–hydroxychloroquine group, and 98 in the placebo group). The SF-36 physical-component summary score did not differ significantly among the three study groups at the end of the treatment period, with mean scores of 35.0 (95% confidence interval [CI], 33.5 to 36.5) in the doxycycline group, 35.6 (95% CI, 34.2 to 37.1) in the clarithromycin–hydroxychloroquine group, and 34.8 (95% CI, 33.4 to 36.2) in the placebo group ($P=0.69$; a difference of 0.2 [95% CI, –2.4 to 2.8] in the doxycycline group vs. the placebo group and a difference of 0.9 [95% CI, –1.6 to 3.3] in the clarithromycin–hydroxychloroquine group vs. the placebo group); the score also did not differ significantly among the groups at subsequent study visits ($P=0.35$). In all study groups, the SF-36 physical-component summary score increased significantly from baseline to the end of the treatment period ($P<0.001$). The rates of adverse events were similar among the study groups. Four serious adverse events thought to be related to drug use occurred during the 2-week open-label ceftriaxone phase, and no serious drug-related adverse event occurred during the 12-week randomized phase.

CONCLUSIONS

In patients with persistent symptoms attributed to Lyme disease, longer-term antibiotic treatment did not have additional beneficial effects on health-related quality of life beyond those with shorter-term treatment. (Funded by the Netherlands Organization for Health Research and Development ZonMw; PLEASE ClinicalTrials.gov number, NCT01207739.)

From the Departments of Medicine (A.B., H.J.M.H., F.J.V., M.L.V., M.T., B.J.K.), Medical Psychology (H.M., M.L.V., A.W.M.E.), Rheumatology (F.H.), and Health Evidence (A.R.T.D.) and the Radboud Center for Infectious Diseases (A.B., H.J.M.H., F.J.V., M.L.V., M.T., B.J.K.), Radboud University Medical Center, and Sint Maartenskliniek (F.J.V., F.H.), — both in Nijmegen, and the Health, Medical, and Neuropsychology Unit, Institute of Psychology, Leiden University, Leiden (H.M., A.W.M.E.) — all in the Netherlands. Address reprint requests to Dr. Kullberg at the Department of Medicine 463, Radboud University Medical Center, P.O. Box 9101, 6500 HB, Nijmegen, the Netherlands.

N Engl J Med 2016;374:1209-20.
DOI: 10.1056/NEJMoa1505425

Patients with Lyme disease, which is caused by the *Borrelia burgdorferi* sensu lato complex (including *B. afzelii* and *B. garinii* in Europe), often report persistent symptoms.[1] These symptoms are also referred to as the post–Lyme disease syndrome or chronic Lyme disease and may occur after resolution of an erythema migrans rash or after other — possibly unnoticed — manifestations of early Lyme disease, regardless of whether a patient received initial appropriate antibiotic treatment. Patients present mainly with pain, fatigue, and neurologic or cognitive disturbances.[2,3]

A Quick Take is available at NEJM.org

Previous randomized, clinical trials have not shown convincingly that prolonged antibiotic treatment has beneficial effects in patients with persistent symptoms attributed to Lyme disease.[4-6] Nonetheless, the debate about this issue has continued.[7] Although most guidelines do not recommend antimicrobial therapy for longer than 2 to 4 weeks,[8,9] other guidelines recommend prolonged antibiotic therapy.[10] We performed a randomized, double-blind, clinical trial (Persistent Lyme Empiric Antibiotic Study Europe [PLEASE]) that included three study groups to compare shorter-term treatment (ceftriaxone followed by placebo [placebo group]) with longer-term treatment (ceftriaxone followed by doxycycline [doxycycline group] or ceftriaxone followed by the combination of clarithromycin and hydroxychloroquine [clarithromycin–hydroxychloroquine group]).

Methods

Study Oversight

The trial was approved by the medical ethics review committee Commissie Mensgebonden Onderzoek regio Arnhem–Nijmegen. The study was conducted in accordance with the principles of the most recent version of the Declaration of Helsinki and the International Conference on Harmonisation guidelines on Good Clinical Practice. Written informed consent was provided by all the participants. All the authors take responsibility for the accuracy and completeness of the reported data and vouch for the fidelity of the trial to the protocol (available with the full text of this article at NEJM.org) and statistical analysis plan (which is included in the protocol). Details of the protocol and study design have been published previously.[11] The trial was performed at two sites in the Netherlands (Radboud University Medical Center and Sint Maartenskliniek) and was overseen by an independent external data and safety monitoring board.

Study Population

Patients were recruited from October 2010 through June 2013. Eligibility was assessed according to previously described inclusion and exclusion criteria (Table S1 in the Supplementary Appendix, available at NEJM.org).[11] In short, patients with persistent symptoms attributed to Lyme disease (musculoskeletal pain, arthritis, arthralgia, neuralgia, sensory disturbances, dysesthesia, neuropsychological disorders, or cognitive disorders, with or without persistent fatigue) were eligible if these symptoms either were temporally related to an erythema migrans rash or an otherwise proven case of symptomatic Lyme disease or were accompanied by *B. burgdorferi* IgG or IgM antibodies, as confirmed by means of immunoblot assay.

Randomization and Blinding

Patients were randomly assigned to one of three groups in a 1:1:1 ratio. Randomization was computerized and balanced by minimization for age (<40 or ≥40 years), sex, duration of symptoms (<1 or ≥1 year), and baseline Global Health Composite score of the RAND-36 Health Status Inventory (RAND SF-36).[12] The randomization list consisted of consecutive medication numbers entered into a secured Web-based database by an independent Web manager. All personnel involved in the study (except the Web manager and study pharmacist) and all participants were unaware of the study-group assignments.

Intervention

All the patients received treatment with 2000 mg of open-label intravenous ceftriaxone daily for 14 days. Patients were admitted at the study site for ceftriaxone administration during days 1 and 2; subsequent doses were given intravenously by specialized home-care nurses. After the 2-week course of ceftriaxone treatment was completed, the patients received a 12-week oral course of doxycycline (100 mg of doxycycline twice daily combined with a placebo twice daily), clarithromycin–hydroxychloroquine (500 mg of clarithromycin twice daily combined with 200 mg of hydroxychloroquine twice daily), or placebo (two

different placebo capsules twice daily), as randomly assigned in a blinded manner. The study drugs and placebo were prepared as capsules with an identical appearance. Active drugs were purchased as standard tablets through the hospital pharmacy department and were placed inside size 000 capsules; placebos were prepared by filling color-matched size 000 capsules with inactive microcrystalline cellulose. Adherence was verified by means of pill counts, patient diaries, and the Medication Event Monitoring System (AARDEX Group), in which microprocessors in the cap of a medication bottle electronically record each time a bottle is opened.[13] The use of specific concomitant medications was prohibited during the entire study period, as described previously.[11]

OUTCOME MEASURES

Outcomes were assessed with the use of self-completed questionnaires at baseline, at the end of the treatment period at 14 weeks (i.e., when the 2-week course of ceftriaxone and the 12-week randomized phase had been completed), at 26 weeks (12 weeks after the end of the treatment period), at 40 weeks (the end of the trial, 26 weeks after the end of the treatment period), and at 52 weeks after the start of the treatment period. Study visits to evaluate safety were scheduled at weeks 2, 8, and 14 and included a medical history, physical examination, and laboratory investigations. The primary outcome measure was health-related quality of life at the end of the treatment period, as assessed by the physical-component summary score of the RAND SF-36.[12,14] This score is based on the weighted T-scores of the four physical scales of the RAND SF-36 (physical functioning, role limitations due to physical health problems, pain, and general health perceptions). The raw SF-36 physical-component summary score was transformed into a norm-based T-score (range, 15 to 61), with a mean (±SD) score of 50±10 in the general population (higher scores indicate a better physical quality of life).

Main secondary outcomes were physical and mental aspects of health-related quality of life, as assessed with the use of the RAND SF-36,[11] and fatigue, as assessed with the use of the fatigue-severity scale of the Checklist Individual Strength, on which scores range from 8 to 56, with higher scores indicating more fatigue[15] (Table 1).

STATISTICAL ANALYSIS

The primary analyses were performed in the modified intention-to-treat population, which included all patients who were randomly assigned to a study group and received at least one dose of ceftriaxone. In the primary analysis, the three study groups were compared at end of the treatment period by means of analysis of covariance, with sex and baseline SF-36 physical-component summary score as covariates. Missing data were imputed according to the baseline-value-carried-forward method. In secondary analyses, linear mixed models were used to evaluate the duration of the treatment effect in an explorative way, and missing data were imputed with the nearest available observation. All models included the baseline value of the dependent variable, sex, time, study-group assignment, and time-by-treatment interaction. No interim efficacy analysis was performed. Sensitivity analyses included a prespecified per-protocol analysis and alternative imputation techniques. Patients who had major protocol violations, such as receipt of less than 75% of a study drug or placebo, as recorded by microprocessors in the Medication Event Monitoring System caps, or use of prohibited concomitant medication, were excluded from the per-protocol analysis.[11]

A two-sided alpha level of 5% was used to indicate statistical significance, and confidence intervals, when calculated, were 95% confidence intervals. Bonferroni correction was used for pairwise comparisons among the three study groups. Statistical analyses were performed with the use of SPSS software, version 20 (SPSS).

The calculation of power was based on a pilot study that included 80 patients with persistent symptoms attributed to Lyme disease.[11] Patients were classified as having a poor or reasonable clinical condition, as assessed during the first clinical consultation at the outpatient clinic. The difference in SF-36 physical-component summary score between patients with a poor clinical condition and those with a reasonable clinical condition was a mean of 3±8 points, which corresponds to the minimal clinically important difference of 2 to 5 points that has been proposed for the SF-36 physical-component summary score.[14] We calculated that a minimum of 75 patients would need to be assigned to each group (225 patients in total) for the study to have 90% power to detect a difference of 3 points at

Table 1. Baseline Characteristics in the Modified Intention-to-Treat Population.*

Characteristic	Doxycycline Group (N=86)	Clarithromycin–Hydroxychloroquine Group (N=96)	Placebo Group (N=98)
Female sex — no. (%)	40 (47)	42 (44)	47 (48)
Age — yr	48.1±12.8	48.2±13.0	50.0±9.7
White race — no. (%)†	84 (98)	96 (100)	98 (100)
Current symptoms — no. (%)‡			
Arthralgia	80 (93)	87 (91)	84 (86)
Musculoskeletal pain	72 (84)	77 (80)	76 (78)
Sensory disturbances	62 (72)	72 (75)	79 (81)
Neuralgia	7 (8)	16 (17)	18 (18)
Neurocognitive symptoms	76 (88)	81 (84)	85 (87)
Fatigue	84 (98)	91 (95)	92 (94)
Duration of symptoms — yr			
Median	2.7	2.7	2.1
Interquartile range	1.3–7.7	1.3–5.4	0.9–5.5
Lyme disease history — no. (%)‡			
Tick bite	47 (55)	46 (48)	60 (61)
Erythema migrans§	25 (29)	26 (27)	27 (28)
Acrodermatitis chronica atrophicans¶	0	1 (1)	2 (2)
Meningoradiculitis‖	1 (1)	9 (9)	5 (5)
Previous antibiotic treatment — no. (%)	75 (87)	86 (90)	89 (91)
Duration — days			
Median	40	30	31
Interquartile range	27–57	21–44	28–58
No. of courses			
Median	2.0	2.0	2.0
Interquartile range	1.0–2.0	1.0–2.0	1.0–2.5
Intravenous treatment — no. (%)	11 (13)	16 (17)	15 (15)
Positive *Borrelia burgdorferi* serology — no. (%)	70 (81)	73 (76)	75 (77)
IgM	25 (29)	21 (22)	35 (36)
IgG	55 (64)	65 (68)	58 (59)
RAND SF-36 score**			
Physical-component summary	30.3±6.3	32.7±7.5	31.8±8.1
Mental-component summary	37.4±9.9	37.1±9.8	37.6±9.6
Global-health composite	32.1±8.0	33.1±8.3	33.0±9.1
Physical-functioning scale	37.3±8.2	40.3±9.9	38.1±9.4
Role–physical scale	28.8±5.9	31.3±9.5	30.3±8.6
Bodily pain scale	35.2±8.3	37.3±8.2	38.1±9.4
General-health scale	35.5±7.7	35.9±7.6	35.9±8.4
Mental-health scale	44.2±9.8	43.6±10.0	44.0±8.5
Role–emotional scale	41.8±15.1	39.9±15.2	42.4±14.8
Social-functioning scale	33.5±12.8	33.8±12.0	34.2±12.2
Vitality scale	38.3±7.1	39.0±7.8	38.3±7.7

Table 1. (Continued.)

Characteristic	Doxycycline Group (N=86)	Clarithromycin–Hydroxychloroquine Group (N=96)	Placebo Group (N=98)
Checklist Individual Strength††			
Total score	101.9±19.4	96.5±20.7	99.3±22.3
Fatigue-severity scale	46.0±8.1	42.7±10.7	43.8±10.6

* Plus–minus values are means ±SD. All study groups received a 2-week course of ceftriaxone before the randomized 12-week course of study drug or placebo. The modified intention-to-treat population included all patients who were randomly assigned to a study group and received at least one dose of ceftriaxone. Between-group differences in characteristics were analyzed with the use of analysis of variance for continuous variables, chi-square tests for proportions, and Fisher's exact test for small numbers (expected frequency <5). Data that were not normally distributed were analyzed with the use of Kruskal–Wallis tests. There were no significant baseline differences among the study groups at a significance level of 0.05. RAND SF-36 denotes the RAND-36 Health Status Inventory.
† Race was self-reported.
‡ Categories are not mutually exclusive.
§ The condition was considered to be temporally related if it was diagnosed by a physician 0 to 4 months before the onset of symptoms.
¶ This condition was considered to be temporally related if it was diagnosed by a physician or biopsy 0 to 4 months before the onset of symptoms.
‖ The condition was considered to be temporally related if it was diagnosed on the basis of intrathecal borrelia antibody production 0 to 4 months before the onset of symptoms.
** The ranges of the RAND SF-36 scores were as follows: physical-component summary, 15 to 61; mental-component summary, 11 to 66; global-health composite, 8 to 65; physical-functioning scale, 16 to 58; role–physical scale, 26 to 56; bodily pain scale, 20 to 60; general-health scale, 20 to 64; mental-health scale, 16 to 66; role–emotional scale, 19 to 54; social-functioning scale, 12 to 57; and vitality scale, 26 to 70. For all scales, higher scores indicate better quality of life.
†† Scores on the Checklist Individual Strength range from 20 to 140 for the total score and from 8 to 56 for the fatigue-severity scale. For both scales, higher scores indicate more fatigue.

a two-sided alpha level of 5% and a reliability coefficient (correlation between consecutive measurements) of 0.7.[16] To compensate for possible loss to follow-up, a study population of at least 255 patients was targeted.

RESULTS

STUDY POPULATION AND BASELINE CHARACTERISTICS

Approximately 1200 patients were screened. The most frequent reasons for ineligibility were negative serologic findings combined with Lyme disease that was either unproven or temporally unrelated to symptoms, a coexisting condition that could account for the patient's symptoms, or known unacceptable side effects from the active study drugs. Of all eligible patients, fewer than 10% declined to participate. A total of 281 patients underwent randomization, and 280 started the oral course of the study drug or placebo (Fig. 1). Table 1 shows the baseline characteristics of patients included in the modified intention-to-treat analysis; there were no significant baseline differences among the study groups. The randomized oral regimen (active study drug or placebo) was completed by 252 patients (90.0%): 76 of 86 patients (88.4%) in the doxycycline group, 84 of 96 patients (87.5%) in the clarithromycin–hydroxychloroquine group, and 92 of 98 patients (93.9%) in the placebo group (P=0.28) (Fig. 1).

No differences in adherence were recorded among the study groups (P=0.50); 75 patients (87.2%) in the doxycycline group, 78 (81.3%) in the clarithromycin–hydroxychloroquine group, and 84 (85.7%) in the placebo group took at least 75% of the assigned study medication or placebo, as recorded by the microprocessors on the Medication Event Monitoring System caps (Fig. 1).

OUTCOMES

The primary outcome in the modified intention-to-treat analysis (i.e., the mean health-related quality of life at the end of the treatment period, as indicated by the SF-36 physical-component summary score, corrected for baseline SF-36 physical-component summary score and sex) did not differ significantly among the study groups (P=0.69) (Table 2). With respect to the secondary outcomes, the mean SF-36 physical-component summary score among all patients in the

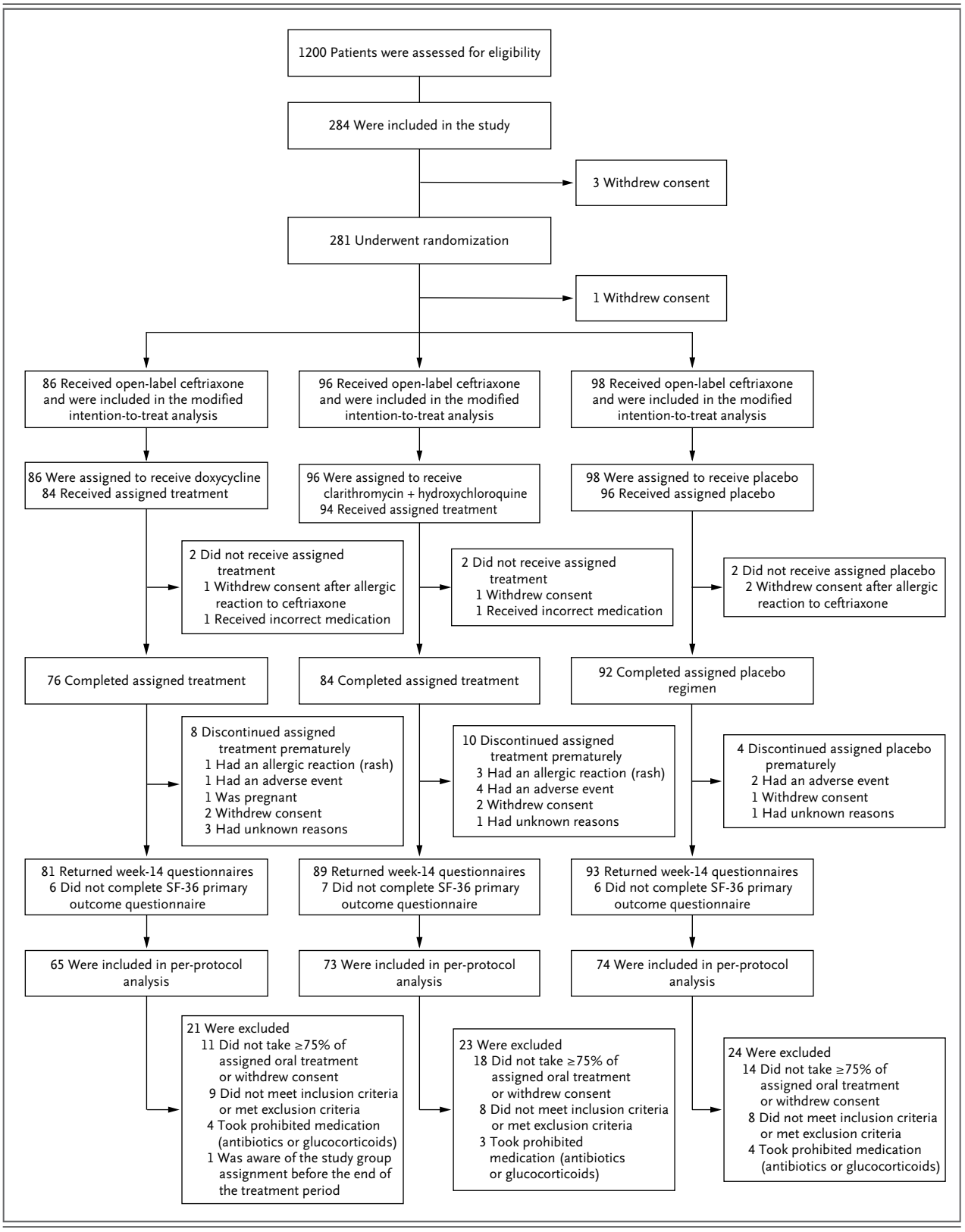

Figure 1 (facing page). Enrollment, Randomization, and Analysis.

Some patients were excluded from the per-protocol analysis because of two or more reasons. Premature discontinuation was defined as discontinuation of the study drug or placebo 7 days or more before the scheduled end of the treatment period, as recorded by microprocessors in the Medication Event Monitoring System caps that were used to track adherence. Week 14 was the end of the treatment period, after the 2-week course of ceftriaxone and the 12-week course of the randomized study drug or placebo had been completed. SF-36 denotes RAND-36 Health Status Inventory.

modified intention-to-treat analysis increased from 31.8 at baseline to 36.4 at the end of the treatment period (difference, 4.6 points; 95% confidence interval [CI], 3.6 to 5.5; P<0.001). At weeks 26, 40, and 52, the SF-36 physical-component summary score remained higher than the baseline score but did not change significantly from the score at the end of the treatment period in any of the study groups (Fig. 2). None of the secondary outcome measures at the end of the treatment period differed significantly among the study groups (Table 2). Mixed-model analyses did not show any additional longer-term treatment effect with respect to the SF-36 physical-component summary score or any of the secondary outcomes; P values for time-by-treatment interaction ranged from 0.14 to 0.90, and there was no significant difference among the study groups in the SF-36 physical-component summary score (P=0.35) or any other secondary outcome measure at any time point during follow-up. All sensitivity analyses yielded results similar to those of the main analyses. Specifically, the results were not quantitatively different when alternate imputation techniques were used for missing data (Table S4 in the Supplementary Appendix). The per-protocol analysis, which included 212 patients (Fig. 1), yielded similar results to the modified intention-to-treat analysis at the end of the treatment period and during follow-up across the three study groups.

SAFETY

Overall, 205 patients (73.2%) reported at least one adverse event, 9 patients (3.2%) had a serious adverse event, and 19 patients (6.8%) had an adverse event that led to discontinuation of the study drug (Table 3). Most adverse events were grade 1 or 2 according to the criteria of the AIDS Clinical Trials Group for grading the severity of adverse events among adults (Table S3 in the Supplementary Appendix).

During the 2-week open-label ceftriaxone phase, 131 patients (46.8%) reported at least one adverse event. Most of these adverse events were judged to be drug-related, and rash and diarrhea were the most common events. No catheter-associated infections were reported. In 6 patients, an allergic adverse event led to the discontinuation of ceftriaxone. Five serious adverse events were reported, four of which were allergic reactions related to ceftriaxone use.

During the 12-week randomized phase, 134 patients (47.9%) had at least one adverse event (Table 3), most of which were judged to be drug-related. The percentage of patients with adverse events from any cause and with drug-related adverse events did not differ significantly among the study groups (P=0.27 and P=0.14, respectively). Photosensitivity and nausea were the most common events in the doxycycline group. Nausea and diarrhea were the most common events in the clarithromycin–hydroxychloroquine group, and rash was significantly more prevalent in that group than in either of the other two groups (P=0.01). Fourteen patients (5.0%) discontinued the randomized active drug or placebo because of an adverse event; the number of patients who discontinued their assigned regimen did not differ significantly among the three study groups (P=0.49). Four serious adverse events were reported, none of which were drug-related.

DISCUSSION

In this randomized, double-blind trial involving patients with persistent symptoms attributed to Lyme disease, prolonged antibiotic treatment (ceftriaxone followed by 12 weeks of either doxycycline or clarithromycin–hydroxychloroquine) did not lead to a better health-related quality of life than that with shorter-term treatment (ceftriaxone followed by placebo). Patients with persistent symptoms attributed to Lyme disease have a poor quality of life, as has been reported in previous studies[5,6,17,18]; the low baseline RAND SF-36 scores of the patients in our trial also reflect the poor quality of life among

Table 2. Treatment Effect at the End of the Treatment Period in the Modified Intention-to-Treat Population.*

Outcome	Doxycycline Group (N = 86)	Clarithromycin–Hydroxychloroquine Group (N = 96)	Placebo Group (N = 98)	P Value†	Doxycycline Group vs. Placebo Group	Clarithromycin–Hydroxychloroquine Group vs. Placebo Group
	score (95% CI)				*difference in score (95% CI)‡*	
Primary outcome: SF-36 physical-component summary§	35.0 (33.5 to 36.5)	35.6 (34.2 to 37.1)	34.8 (33.4 to 36.2)	0.69	0.2 (–2.4 to 2.8)	0.9 (–1.6 to 3.3)
Secondary outcomes						
RAND SF-36§						
Mental-component summary	40.2 (38.6 to 41.9)	40.5 (38.9 to 42.1)	40.1 (38.6 to 41.7)	0.94	0.1 (–2.7 to 2.9)	0.4 (–2.3 to 3.1)
Global-health composite	36.1 (34.5 to 37.8)	36.6 (35.1 to 38.1)	36.0 (34.5 to 37.5)	0.85	0.1 (–2.6 to 2.9)	0.6 (–2.1 to 3.2)
Physical-functioning scale	41.9 (40.5 to 43.3)	42.1 (40.8 to 43.4)	41.0 (39.7 to 42.3)	0.44	0.9 (–1.4 to 3.2)	1.1 (–1.1 to 3.4)
Role–physical scale	33.6 (31.6 to 35.6)	34.4 (32.5 to 36.3)	33.9 (32.0 to 35.8)	0.84	–0.3 (–3.7 to 3.1)	0.5 (–2.8 to 3.8)
Bodily pain scale	39.1 (37.5 to 40.7)	40.5 (39.0 to 41.9)	39.4 (37.9 to 40.9)	0.42	–0.3 (–2.9 to 2.4)	1.1 (–1.5 to 3.6)
General-health scale	37.1 (35.6 to 38.6)	38.4 (37.0 to 39.8)	37.5 (36.2 to 38.9)	0.41	–0.4 (–2.9 to 2.0)	0.9 (–1.5 to 3.3)
Mental-health scale	45.1 (43.8 to 46.4)	45.2 (43.9 to 46.4)	45.1 (43.9 to 46.4)	1.00	0.0 (–2.3 to 2.2)	0.0 (–2.1 to 2.2)
Role–emotional scale	44.7 (42.4 to 47.0)	41.4 (39.2 to 43.6)	42.6 (40.4 to 44.8)	0.11	2.1 (–1.7 to 6.0)	–1.2 (–5.0 to 2.6)
Social-functioning scale	36.3 (34.2 to 38.4)	38.5 (36.6 to 40.5)	37.5 (35.6 to 39.5)	0.32	–1.2 (–4.7 to 2.3)	1.0 (–2.4 to 4.4)
Vitality scale	42.5 (40.9 to 44.0)	42.4 (41.0 to 43.9)	41.9 (40.5 to 43.4)	0.85	0.5 (–2.0 to 3.1)	0.5 (–2.0 to 3.0)
Checklist Individual Strength¶						
Total score	88.7 (84.4 to 92.9)	87.1 (83.0 to 91.1)	88.4 (84.4 to 92.4)	0.84	0.3 (–6.9 to 7.4)	–1.3 (–8.3 to 5.6)
Fatigue-severity scale	39.4 (37.3 to 41.5)	38.6 (36.6 to 40.5)	38.3 (36.3 to 40.2)	0.73	1.1 (–2.4 to 4.6)	0.3 (–3.1 to 3.7)

* All study groups first received a 2-week course of ceftriaxone before the randomized 12-week course of study drug or placebo. P values were derived by analysis of covariance. All scores are adjusted for sex and baseline SF-36 physical-component summary score.
† Bonferroni correction was used for pairwise comparisons among the three study groups.
‡ Group differences should exceed 2 to 4 T-points (exact number of points varies for each scale) to indicate minimally important differences on all RAND SF-36 scales.[14]
§ The ranges of the RAND SF-36 scores were as follows: RAND SF-36 physical-component summary, 15 to 61; mental-component summary, 11 to 66; global-health composite, 8 to 65; physical-functioning scale, 16 to 58; role–physical scale, 26 to 56; bodily pain scale, 20 to 60; general-health scale, 20 to 64; mental-health scale, 16 to 66; role–emotional scale, 19 to 54; social-functioning scale, 12 to 57; and vitality scale, 26 to 70. For all scales, higher scores indicate better quality of life.
¶ Scores on the Checklist Individual Strength range from 20 to 140 for the total score and from 8 to 56 for the fatigue-severity scale. For both scales, higher scores indicate more fatigue.

these patients. At the 14-week visit at the end of the treatment period, the mean SF-36 physical-component summary score had improved significantly from baseline regardless of the study-group assignment, but quality of life remained below that of the general population. Similar improvements over time, regardless of study-group assignment, were reported by Kaplan et al., who compared placebo with ceftriaxone followed by doxycycline for persistent symptoms attributed to Lyme disease.[19]

Whether improvement in the SF-36 physical-component summary score at the end of the treatment period is a beneficial effect of shorter-term antibiotic therapy or a nonspecific effect caused by the low level of baseline functioning, expectations associated with treatment, or placebo effects remains unclear, because all the patients had received 2 weeks of open-label antibiotics before entering into the longer-term randomized study-drug or placebo phase. No significant differences among the study groups were found for any of the secondary outcomes at the end of the treatment period. In addition, no significant changes over time were observed during the 26-week follow-up after the end of the treatment period in any of the study groups.

Although we did not find a significant benefit of longer-term antibiotic therapy, we did find that there were side effects from the use of antibiotics; however, these side effects were similar among the study groups. The majority of patients (68.6%) reported a drug-related adverse event. During the open-label ceftriaxone phase, the incidence of serious adverse events was low; no patient had a serious adverse event related to the use of catheters, and 4 of 280 patients (1.4%) had allergic reactions. During the randomized phase, photosensitivity related to doxycycline use and rash related to clarithromycin–hydroxychloroquine use were the most common adverse events, and no serious adverse event was thought to be related to the randomized study drugs or placebo.

Specific efforts were made to ensure that the patients adhered to the study regimens. Using the Medication Event Monitoring System caps, we recorded that 22 patients (7.9%) discontinued treatment 7 days or more before the end of the treatment period at week 14. In a sensitivity analysis that included the 212 patients who were more than 75% adherent to the study regimen, as determined by electronic medication bottle caps, and had no major protocol violations, no significant difference was shown among the study groups.

The findings of the current trial contribute to the findings of prior work.[4-6,18] Our results are consistent with those from the randomized, placebo-controlled trials by Klempner et al.,[5] who did not identify a benefit from treatment with ceftriaxone followed by doxycycline for a total of 90 days. However, these trials had been performed in North America, and Lyme disease in Europe is caused by different borrelia species.[20] The trials by Klempner et al.[5] have been the subject of divergent opinions because they were discontinued prematurely after an interim analysis had indicated that a significant difference in efficacy was unlikely to be reached. Therefore, although the results are statistically

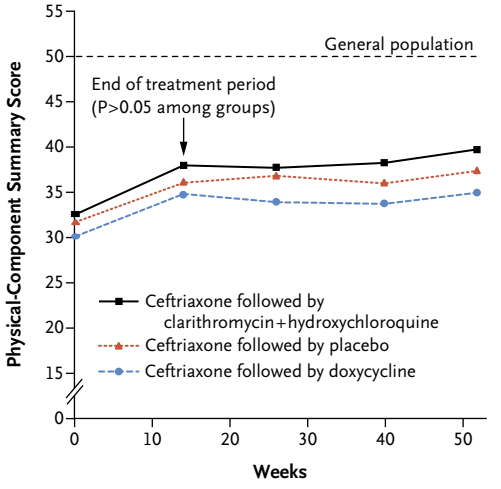

Figure 2. Physical-Component Summary Scores.
Shown is the mean SF-36 physical-component summary score for each study group at baseline and at subsequent study visits (nonimputed values). The SF-36 physical-component summary score is based on the weighted T-scores of the four physical RAND SF-36 scales (physical functioning, role limitations due to physical health problems, pain, and general health perceptions). The raw SF-36 physical-component summary score was transformed into a norm-based T-score (range, 15 to 61), with a mean (±SD) score of 50±10 in the general population (higher scores indicate a better physical quality of life). The P value was derived by means of analysis of covariance at the end of the treatment period at 14 weeks, with adjustment for sex and baseline SF-36 physical-component summary score.

Table 3. Adverse Events in the Modified Intention-to-Treat Population.*

Type of Event	Total (N=280)	Open-Label Phase (N=280)	Randomized Phase			P Value
			Doxycycline Group (N=86)	Clarithromycin–Hydroxychloroquine Group (N=96)	Placebo Group (N=98)	
			no. of participants (percent)			
Any adverse event†	205 (73.2)	131 (46.8)	47 (54.7)	45 (46.9)	42 (42.9)	0.27
Any drug-related adverse event†	192 (68.6)	127 (45.4)	42 (48.8)	42 (43.8)	34 (34.7)	0.14
Discontinued treatment owing to adverse event†	19 (6.8)	6 (2.1)	3 (3.5)	7 (7.3)	4 (4.1)	0.49‡
Any serious adverse event	9 (3.2)	5 (1.8)	3 (3.5)	1 (1.0)	0	0.08‡
Most common adverse events						
Diarrhea	91 (32.5)	72 (25.7)	4 (4.7)	9 (9.4)	6 (6.1)	0.43
Nausea	44 (15.7)	20 (7.1)	9 (10.5)	10 (10.4)	5 (5.1)	0.31
Rash†	31 (11.1)	23 (8.2)	1 (1.2)	8 (8.3)	1 (1.0)	0.01‡
Mucosal fungal infection	20 (7.1)	8 (2.9)	5 (5.8)	4 (4.2)	3 (3.1)	0.66‡
Photosensitivity	19 (6.8)	2 (0.7)	16 (18.6)	0	1 (1.0)	<0.001
Headache	16 (5.7)	12 (4.3)	0	2 (2.1)	2 (2.0)	0.55‡
Dizziness	16 (5.7)	3 (1.1)	3 (3.5)	5 (5.2)	5 (5.1)	0.88‡
Visual impairment	16 (5.7)	1 (0.4)	1 (1.2)	4 (4.2)	10 (10.2)	0.02‡

* Data are the number of patients who had at least one event of a given type (% of study group). All patients received a 2-week course of ceftriaxone treatment (open-label phase), after which patients were randomly assigned to receive a 12-week oral course of doxycycline, clarithromycin–hydroxychloroquine, or placebo (randomized phase).

† The total is not a sum of the two trial phases because some patients had an adverse event during both phases. P values were derived from the chi-square test for the comparisons of the three study groups during the randomized phase.

‡ Fisher's exact test was used when the numbers were small (expected frequency <5).

valid, the value of prolonged antibiotic therapy for patients with Lyme disease has been based on a study population of approximately 115 patients. Others have suggested that the trials by Klempner et al. were underpowered as a result of an optimistic estimate of the size of the treatment effect.[7] In a pilot study, we determined that the clinically relevant treatment effect on the SF-36 physical-component summary score was 3 points, as was recommended by the SF-36 Health Survey.[14] None of the differences among the study groups were found to exceed the minimal clinically relevant difference for each of the RAND SF-36 scales, which varies between 2 and 4 across scales.[14] Whereas earlier trials might have been influenced by baseline differences, we included baseline health-related quality of life as a covariate.

Three other small, placebo-controlled trials have addressed prolonged treatment for persistent symptoms attributed to Lyme disease and showed positive effects for some outcomes only.[4,6,18] Krupp et al.[4] reported a significant treatment effect of ceftriaxone on fatigue, but not on cognitive function, at follow-up. Fallon et al. found a beneficial effect of ceftriaxone on neurocognitive performance at week 12, but the effect was not sustained to week 24.[18] Cameron et al. reported beneficial effects of amoxicillin on mental-health scores, but not on physical health, in a subgroup of patients.[6] Although several noncomparative, open-label studies have shown beneficial effects of prolonged antimicrobial treatment, including the regimens used in the current study,[21-24] randomized, controlled trials of prolonged antimicrobial treatment have not confirmed those effects.

The current trial has several limitations.

First, patients received open-label antibiotics for 2 weeks before the randomized phase. Consequently, the study was designed to compare longer-term therapy with shorter-term therapy, rather than with placebo as was done in previous trials.[4,5,18] Although we did not identify any benefit of longer-term therapy, the question of whether a 2-week regimen of antibiotics is superior to withholding any therapy in our patient population remains unanswered. We chose not to include a study group that received only placebo because it was judged to be unethical to withhold treatment from patients who might have an infection at baseline that had not yet been treated. We selected ceftriaxone because it is considered the treatment of choice for disseminated Lyme disease.[5,8] Thus, although 14 weeks of antimicrobial therapy did not provide a clinical benefit for patients with persistent symptoms attributed to Lyme disease, our results cannot show whether our study may have included patients with undiagnosed active *B. burgdorferi* infection, who have benefited from ceftriaxone treatment.

This trial, as well as previous trials,[4-6,18] was aimed at the treatment of patients with persistent, notably distressing or impairing symptoms that emerged after well-documented Lyme disease. We acknowledge that the cause of these persistent symptoms is unclear and that these patients may be heterogeneous with respect to the pathogenesis or the duration and severity of the symptoms — which reflects the heterogeneity of the population seen in clinical practice. We prevented an imbalance in baseline characteristics among the study groups by performing a randomization balanced for duration of symptoms (<1 or ≥1 year) and baseline RAND SF-36 score. Finally, it may be argued that 14 weeks of treatment is insufficient to show a beneficial treatment effect. However, whereas prolonged antimicrobial treatment is not uncommon for various infectious diseases,[25,26] the purpose of prolonged therapy for such diseases is for the prevention of microbiologic relapse rather than for a delayed onset of clinical alleviation of signs or symptoms. We are not aware of any infectious disease in which the initial effect on signs, symptoms, and laboratory findings is delayed beyond the first 3 months of effective therapy.

In conclusion, the current trial suggests that 14 weeks of antimicrobial therapy does not provide clinical benefit beyond that with shorter-term treatment among patients who present with fatigue or musculoskeletal, neuropsychological, or cognitive disorders that are temporally related to prior Lyme disease or accompanied by positive *B. burgdorferi* serologic findings.

Supported by a grant (171002304) from the Netherlands Organization for Health Research and Development ZonMw.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the data and safety monitoring board members, George E. Griffin, M.D. (University of London), Joanna in 't Hout, M.Sc. (Radboud University, Nijmegen), and Peterhans J. van den Broek, M.D. (Leiden University), as well as Stephan Keijmel, M.D., Daniëlle van den Berg, M.D., Hans Groenewoud, M.Sc., Angela Colbers, M.Sc., Hanneke Geurts, M.Sc., and the study pharmacists at Radboud University Medical Center and Sint Maartenskliniek for their assistance.

REFERENCES

1. van den Wijngaard CC, Hofhuis A, Harms MG, et al. The burden of Lyme borreliosis expressed in disability-adjusted life years. Eur J Public Health 2015;25:1071-8.
2. Cairns V, Godwin J. Post-Lyme borreliosis syndrome: a meta-analysis of reported symptoms. Int J Epidemiol 2005;34:1340-5.
3. Shadick NA, Phillips CB, Sangha O, et al. Musculoskeletal and neurologic outcomes in patients with previously treated Lyme disease. Ann Intern Med 1999;131:919-26.
4. Krupp LB, Hyman LG, Grimson R, et al. Study and Treatment of Post Lyme Disease (STOP-LD): a randomized double masked clinical trial. Neurology 2003;60:1923-30.
5. Klempner MS, Hu LT, Evans J, et al. Two controlled trials of antibiotic treatment in patients with persistent symptoms and a history of Lyme disease. N Engl J Med 2001;345:85-92.
6. Cameron D. Severity of Lyme disease with persistent symptoms: insights from a double-blind placebo-controlled clinical trial. Minerva Med 2008;99:489-96.
7. Delong AK, Blossom B, Maloney EL, Phillips SE. Antibiotic retreatment of Lyme disease in patients with persistent symptoms: a biostatistical review of randomized, placebo-controlled, clinical trials. Contemp Clin Trials 2012;33:1132-42.
8. Wormser GP, Dattwyler RJ, Shapiro ED, et al. The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 2006;43:1089-134.
9. Mygland A, Ljøstad U, Fingerle V, Rupprecht T, Schmutzhard E, Steiner I. EFNS guidelines on the diagnosis and management of European Lyme neuroborreliosis. Eur J Neurol 2010;17:8-16.
10. Cameron DJ, Johnson LB, Maloney EL. Evidence assessments and guideline recommendations in Lyme disease: the clinical management of known tick bites, erythema migrans rashes and persistent disease. Expert Rev Anti Infect Ther 2014;12:1103-35.
11. Berende A, ter Hofstede HJ, Donders ART, et al. Persistent Lyme Empiric Antibiotic Study Europe (PLEASE) — design of a randomized controlled trial of pro-

longed antibiotic treatment in patients with persistent symptoms attributed to Lyme borreliosis. BMC Infect Dis 2014;14:543.
12. Hays RD. RAND-36 Health Status Inventory. San Antonio, TX: Psychological Corporation, 1998.
13. Urquhart J. The electronic medication event monitor: lessons for pharmacotherapy. Clin Pharmacokinet 1997;32:345-56.
14. Maruish ME. User's manual for the SF-36v2 Health Survey. Lincoln, RI: Quality Metric, 2011.
15. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res 1994;38:383-92.
16. Borm GF, Fransen J, Lemmens WA. A simple sample size formula for analysis of covariance in randomized clinical trials. J Clin Epidemiol 2007;60:1234-8.
17. Cunningham WE, Nakazono TT, Tsai KL, Hays RD. Do differences in methods for constructing SF-36 physical and mental health summary measures change their associations with chronic medical conditions and utilization? Qual Life Res 2003;12:1029-35.
18. Fallon BA, Keilp JG, Corbera KM, et al. A randomized, placebo-controlled trial of repeated IV antibiotic therapy for Lyme encephalopathy. Neurology 2008;70:992-1003.
19. Kaplan RF, Trevino RP, Johnson GM, et al. Cognitive function in post-treatment Lyme disease: do additional antibiotics help? Neurology 2003;60:1916-22.
20. van Dam AP, Kuiper H, Vos K, et al. Different genospecies of Borrelia burgdorferi are associated with distinct clinical manifestations of Lyme borreliosis. Clin Infect Dis 1993;17:708-17.
21. Clarissou J, Song A, Bernede C, et al. Efficacy of a long-term antibiotic treatment in patients with a chronic tick associated poly-organic syndrome (TAPOS). Med Mal Infect 2009;39:108-15.
22. Donta ST. Tetracycline therapy for chronic Lyme disease. Clin Infect Dis 1997;25:Suppl 1:S52-6.
23. Donta ST. Macrolide therapy of chronic Lyme disease. Med Sci Monit 2003;9:PI136-42.
24. Stricker R, DeLong AK, Green CL, et al. Benefit of intravenous antibiotic therapy in patients referred for treatment of neurologic Lyme disease. Int J Gen Med 2011;4:639-46.
25. Raoult D, Houpikian P, Tissot Dupont H, Riss JM, Arditi-Djiane J, Brouqui P. Treatment of Q fever endocarditis: comparison of 2 regimens containing doxycycline and ofloxacin or hydroxychloroquine. Arch Intern Med 1999;159:167-73.
26. British Thoracic Society. A controlled trial of 6 months' chemotherapy in pulmonary tuberculosis: final report: results during the 36 months after the end of chemotherapy and beyond. Br J Dis Chest 1984;78:330-6.