LETTER

doi:10.1038/nature10384

The evolution of overconfidence

Dominic D. P. Johnson[1] & James H. Fowler[2]

Confidence is an essential ingredient of success in a wide range of domains ranging from job performance and mental health to sports, business and combat[1–4]. Some authors have suggested that not just confidence but overconfidence—believing you are better than you are in reality—is advantageous because it serves to increase ambition, morale, resolve, persistence or the credibility of bluffing, generating a self-fulfilling prophecy in which exaggerated confidence actually increases the probability of success[3–8]. However, overconfidence also leads to faulty assessments, unrealistic expectations and hazardous decisions, so it remains a puzzle how such a false belief could evolve or remain stable in a population of competing strategies that include accurate, unbiased beliefs. Here we present an evolutionary model showing that, counterintuitively, overconfidence maximizes individual fitness and populations tend to become overconfident, as long as benefits from contested resources are sufficiently large compared with the cost of competition. In contrast, unbiased strategies are only stable under limited conditions. The fact that overconfident populations are evolutionarily stable in a wide range of environments may help to explain why overconfidence remains prevalent today, even if it contributes to hubris, market bubbles, financial collapses, policy failures, disasters and costly wars[9–13].

Humans show many psychological biases, but one of the most consistent, powerful and widespread is overconfidence. Most people show a bias towards exaggerated personal qualities and capabilities, an illusion of control over events, and invulnerability to risk (three phenomena collectively known as 'positive illusions')[2–4,14]. Overconfidence amounts to an 'error' of judgement or decision-making, because it leads to overestimating one's capabilities and/or underestimating an opponent, the difficulty of a task, or possible risks. It is therefore no surprise that overconfidence has been blamed throughout history for high-profile disasters such as the First World War, the Vietnam war, the war in Iraq, the 2008 financial crisis and the ill-preparedness for environmental phenomena such as Hurricane Katrina and climate change[9,12,13,15,16].

If overconfidence is both a widespread feature of human psychology and causes costly mistakes, we are faced with an evolutionary puzzle as to why humans should have evolved or maintained such an apparently damaging bias. One possible solution is that overconfidence can actually be advantageous on average (even if costly at times), because it boosts ambition, morale, resolve, persistence or the credibility of bluffing. If such features increased net payoffs in competition or conflict over the course of human evolutionary history, then overconfidence may have been favoured by natural selection[5–8].

However, it is unclear whether such a bias can evolve in realistic competition with alternative strategies. The null hypothesis is that biases would die out, because they lead to faulty assessments and suboptimal behaviour. In fact, a large class of economic models depend on the assumption that biases in beliefs do not exist[17]. Underlying this assumption is the idea that there must be some evolutionary or learning process that causes individuals with correct beliefs to be rewarded (and thus to spread at the expense of individuals with incorrect beliefs). However, unbiased decisions are not necessarily the best strategy for maximizing benefits over costs, especially under conditions of competition, uncertainty and asymmetric costs of different types of error[8,18–21]. Whereas economists tend to posit the notion of human brains as general-purpose utility maximizing machines that evaluate the costs, benefits and probabilities of different options on a case-by-case basis, natural selection may have favoured the development of simple heuristic biases (such as overconfidence) in a given domain because they were more economical, available or faster.

Here we present a model showing that, under plausible conditions for the value of rewards, the cost of conflict, and uncertainty about the capability of competitors, there can be material rewards for holding incorrect beliefs about one's own capability. These adaptive advantages of overconfidence may explain its emergence and spread in humans, other animals or indeed any interacting entities, whether by a process of trial and error, imitation, learning or selection. The situation we model—a competition for resources—is simple but general, thereby capturing the essence of a broad range of competitive interactions including animal conflict, strategic decision-making, market competition, litigation, finance and war.

Suppose a resource r is available to an individual that claims it, and there are two individuals, i and j. These individuals each have initial 'capability' θ_i and θ_j that determine whether or not they would win a conflict over the resource. Without loss of generality, we assume that θ is distributed in the population according to a symmetric stable probability density[22] with cumulative distribution Φ, a mean of 0, and a variance of 0.5. The initial advantage to individual i is $a = \theta_i - \theta_j$, and assumptions about the distribution of θ imply that the probability density of a has a cumulative distribution Φ, a mean of 0, and unit variance (see Supplementary Information for the full model).

If neither individual claims the resource, no fitness is gained. If only one makes a claim, then the claimant acquires the resource and gains fitness r and the other individual gains nothing. If both claim the resource, then both pay a cost c as a result of the conflict between them, but the individual with the higher initial capability will win the conflict, acquiring the resource and obtaining fitness r. This means there are only three outcomes that have an impact on an individual's fitness: winning a conflict (W), losing a conflict (L), and obtaining an unclaimed resource (O). Given the probability of each of these outcomes (p_W, p_L and p_O), the benefits of obtaining the resource r, and the costs of conflict c, the expected fitness is $E(f) = p_W(r - c) + p_L(-c) + p_O(r)$. Note that r and c can denote expected benefits and costs—if conflict outcomes were made probabilistic instead of deterministic, the results would not change.

Individuals choose whether or not to claim a resource on the basis of their perceived capability relative to the capability of other claimants. If there were no uncertainty in this assessment, there would never be a conflict because the dispute can be settled without cost (the stronger individual takes the resource, and the weaker individual surrenders it, allowing both agents to avoid c)[23–26]. In the real world, however, uncertainty is common. We therefore model an individual's uncertainty about his or her opponent's capability by adding an error term v to the opponent's capability such that individual i thinks the capability of individual j is $\theta_j + v_i$. To derive analytical results, we assume that this perception error has a magnitude of $\varepsilon > 0$ and is binomially distributed,

[1]Politics and International Relations, University of Edinburgh, Edinburgh EH8 9LD, UK. [2]Division of Medical Genetics and Department of Political Science, University of California, San Diego, California 92093, USA.

with $\Pr(v = \varepsilon) = \Pr(v = -\varepsilon) = 0.5$ (the 'binomial model'). To evaluate the role of confidence, we allow individuals to perceive their own capability as $\theta + k$, where $k = 0$ indicates unbiased individuals who perceive their capability correctly, $k > 0$ indicates overconfident individuals who think they are stronger than they actually are, and $k < 0$ indicates underconfident individuals who think they are weaker than they actually are.

We explore the emergence and stability of biases in hypothetical populations by using standard assumptions about evolutionary dynamics[27] under which the fittest are more likely to survive or reproduce, or the less fit are more likely to copy better strategies. Figure 1a shows regions of the parameter space and five equilibria that occur in the binomial model, all confirmed both analytically and numerically (see Supplementary Information).

When $r/c > 3/2$, the unique equilibrium is a pure (monomorphic) population of overconfident individuals, all of whom evolve a level of overconfidence that is equal to the size of the perception error ($k^* = \varepsilon$). As long as there is at least some perception error, overconfident individuals resist invasion by all other individuals, including underconfident ($k < 0$), unbiased ($k = 0$) and other kinds of overconfident individuals ($k > 0$).

When $1/3 < r/c < 3/2$, there are two equilibria. First, a mixed (polymorphic) population made up of overconfident individuals ($k^* = \varepsilon$) and underconfident individuals ($k^* = -\varepsilon$) is always possible as long as there is at least some perception error. Second, an unbiased equilibrium ($k^* = 0$) is also possible in this region, but only if the perception error is sufficiently low.

Finally, when $r/c < 1/3$ there are two more equilibria. A pure equilibrium of underconfident individuals ($k^* = -\varepsilon$) is always possible, and a mixed equilibrium of very underconfident ($k^* = -2\varepsilon$) and unbiased ($k^* = 0$) individuals is possible when there is a moderate amount of uncertainty.

The underlying assumptions of the binomial model are deliberately simple to make closed-form characterizations tractable. We also used numerical simulation methods to evaluate the model when we allow the perception error v to vary continuously, using a normal distribution with mean 0 and standard deviation ε (the 'normal model'; Supplementary Information). This assumption may be more realistic than the binomial assumption because it allows perception errors to vary in magnitude.

As with the binomial model, the normal model shows that overconfidence ($k^* > 0$) is the unique pure equilibrium when the benefit/cost ratio is high enough (roughly $r/c > 0.7$; see Fig. 1b), which is notably less stringent than the binomial model reported above. When the benefit/cost ratio falls below this critical value, the unique pure equilibrium is underconfidence ($k^* < 0$). If there is any perception error whatsoever, an absence of bias is only an equilibrium at a single point—the value of resources and the cost of conflict must be in perfect

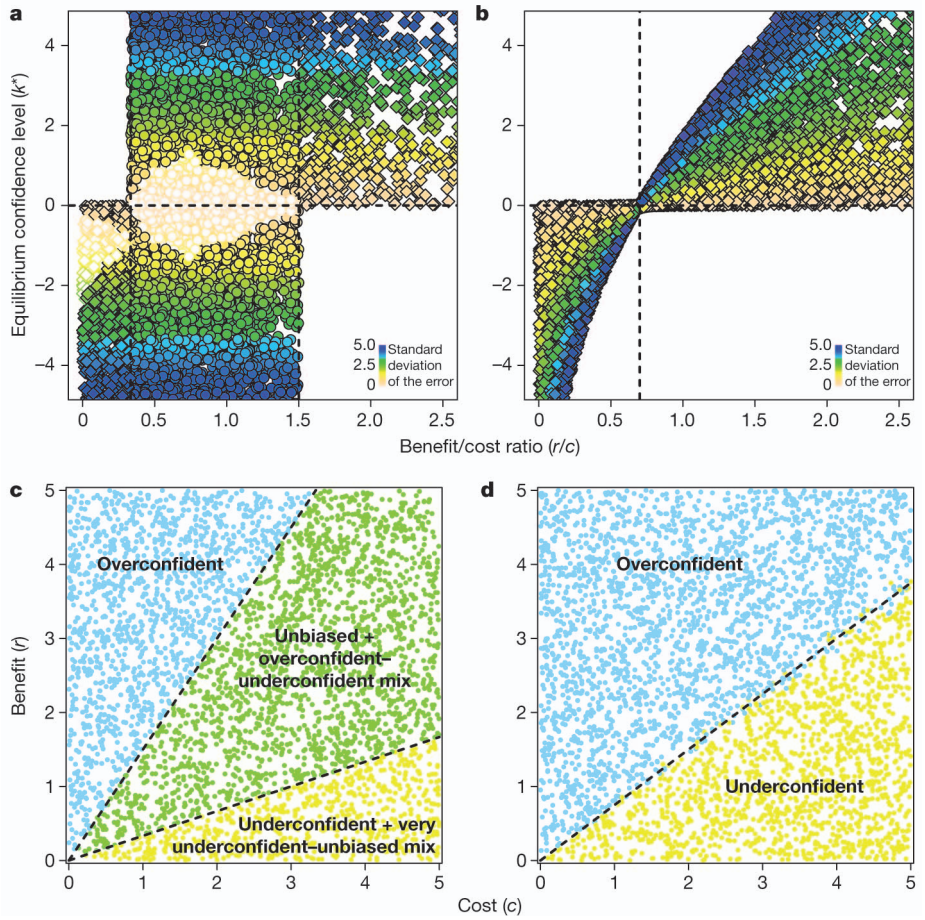

Figure 1 | Best performing levels of confidence across different parameter values. **a**, **b**, Equilibrium levels of confidence k^* for varying benefit/cost ratios (r/c) and degrees of uncertainty about the capabilities of competitors when assessment errors are modelled with a binomial distribution (**a**) or a normal distribution (**b**). Each point shows the results from a single simulation where the cost, benefit and degree of uncertainty were drawn from a uniform distribution (see Supplementary Information). Each panel shows a total of 10,000 simulations. Shapes indicate types of equilibrium that exist for a given parameter combination: diamonds, monomorphic; circles, polymorphic (filled shapes indicate that unbiased strategies are not possible). Colours indicate the degree of uncertainty (the standard deviation of the error as defined on the scales). **c**, **d**, The same results for the binomial (**c**) and normal (**d**) models as a function of costs and benefits (colours indicate what kind of equilibria are possible; these results hold for all levels of perception error). Both models show that overconfident strategies are the unique equilibrium when the benefit/cost ratio is sufficiently high, and unbiased strategies are only possible under limited conditions.

balance to eliminate bias (Fig. 1b). This result suggests that models based on the assumption that individuals perceive their own capabilities without bias[17] are unrealistic: any small change in the benefit/cost ratio will tilt the advantage away from unbiased individuals towards those that assume they are more or less capable than they really are.

The normal model also yields the same positive relationship between perception error and confidence that was derived in the binomial model. As uncertainty about opponent capabilities increases, it becomes more advantageous to express stronger bias (the overconfident become even more confident, and the underconfident become even less confident).

The simulations allowed us to examine some extensions of the model (see Supplementary Information). If we generalize the model to three players, overconfidence is favoured at the same threshold ($r/c > 0.7$). Results are also robust if we allow conflict costs to vary between winners and losers. In fact, the threshold required for overconfidence decreases as losers suffer more. For example, overconfidence evolves when $r/c > 0.6$ if the costs to the winner are $0.8c$, and it evolves when $r/c > 0.45$ if the costs to the winner are $0.2c$. In other words, when conflict for the winner is cheap, overconfidence is even more likely to evolve and persist.

Our model shares interesting parallels with the famous Hawk–Dove game in evolutionary game theory[24]. 'Hawks' escalate until they win (with benefit b) or sustain significant injury (with cost c). 'Doves' only display, and retreat if attacked. Where $b > c$, Hawks take over the population and animals always fight. Where $c > b$, a mixed population of Hawks and Doves emerges. The Hawk–Dove game is important because it shows that (where $c > b$) contests can be resolved by 'conventional' signals (displays only) with minimal fighting—explaining why many animals have dangerous weapons (such as sharp horns or teeth) but death is rare.

We find that the Hawk–Dove game is a special case of our model, in which the only possible strategies are to be infinitely overconfident ($k = \infty$; that is, Hawk) and therefore always claim the resource, or infinitely underconfident ($k = -\infty$; that is, Dove) and therefore never claim. As we show (see Supplementary Information), the standard equilibria of the Hawk–Dove game emerge under these conditions. Strikingly, however, somewhat overconfident (but not infinitely overconfident) individuals always beat both Hawk and Dove. Our model therefore shows that individuals with a more nuanced strategy—even a biased one—do better than the 'extreme' strategies of Hawk and Dove. Moreover, hawkish (overconfident) strategies can dominate even where $c > r$, a finding that contrasts with previous Hawk–Dove models.

Another important result of the model is that environments with more valuable resources will generate more conflict (see Supplementary Information). This parallels the finding in the literature on animal fighting that, where very valuable resources are at stake, hawkish strategies become more common and, in contrast with much animal conflict that is ritualized and restrained, fighting under these conditions can become lethal[28].

The analysis here demonstrates that overconfidence often prevails over accurate assessment. Overconfidence is advantageous because it encourages individuals to claim resources they could not otherwise win if it came to a conflict (stronger but cautious rivals will sometimes fail to make a claim), and it keeps them from walking away from conflicts they would surely win. These results conform with previous observations that systematic overestimates of the probability of winning simple gambling games can be adaptive if the benefits of the resource at stake sufficiently exceed the costs of attempting to gain it[19,20], that aggressive strategies (such as 'Hawk' in Hawk–Dove games) are favoured if the advantages of winning exceed the costs of injury[24], and that overconfident states can outperform others in an agent-based model of conflict[29].

Note that overconfidence in our model is purely self-deception—there is no other-deception ('bluffing') because there is no signalling of k (opponents are not gullible to others' inflated beliefs). This is important because it demonstrates that there are adaptive advantages of overconfidence irrespective of any possible (additional) advantages of bluffing. Bluffing is often argued to be unstable in nature because there would be strong selection on discriminating responses. However, this may be partly why self-deception evolved: "hiding the truth from yourself to hide it more deeply from others"[6,7]. Previous work has also shown that bluffing can survive counter-selection if there is ambiguity in one's own or others' strengths. If so, bluffs and reality cannot be reliably distinguished, and calling another's bluff takes on a cost of its own. It has been suggested[24] that bluffing is therefore more likely (even if it is detectable in principle) among animals in which serious injury is possible—that is, those with weapons—because the costs of calling a bluff can be high.

Our model applies to any replicating entity or any species, but it has particular implications for humans. First, if contested resources were sufficiently valuable compared with the costs of competing for them during human evolutionary history, we might expect humans to have evolved a bias towards overconfidence[5,12,19,20]. Such an outcome is exactly what the literature on experimental psychology has long demonstrated but has lacked an explanation for its origin[2–4,14]. A recent review of whether any 'false beliefs' could be biologically adaptive concluded that there is just a single compelling candidate: positive illusions[8]. Today, we may retain evolved proximate mechanisms that give rise to overconfidence even in situations in which the costs of conflict have increased relative to the value of the reward, making overconfidence maladaptive in many modern settings (such as, perhaps, in interpersonal aggression and war).

Second, overconfidence can arise and spread more quickly among humans than other organisms. Rather than relying on genetic mutation and natural selection over many generations, overconfidence in humans can emerge and spread much more rapidly by other means such as trial and error, imitation or learning (which may also generate considerable variation among different 'ecological' contexts such as habitats, cultures or organizations). These processes of cultural selection may affect how different levels of confidence emerge, survive and spread today among interacting entities, whether individuals, groups, negotiators, lawyers, traders, banks, sports teams, firms, armies or states. In many of these settings, overconfidence may be beneficial on average even though it only attracts attention when it causes costly disasters, or when the environment (the ratio r/c) changes such that overconfidence begins to generate net costs.

Other recent models have explored the evolution of risk preferences[30]; however, in the present model, individuals do not prefer or avoid risk—their heuristic is simply to assess capabilities and claim the resource if they perceive a capability gap. As we show (see Supplementary Information), this heuristic causes individuals to behave as though they were calculating the expected outcome of a risky choice under a specific set of assumptions about themselves and their opponents and comparing it with a required risk premium, which is cognitively a much more demanding task. Thus, although it is possible that risk preferences contribute to behaviour in competition and conflict, the simpler mechanism of overconfidence provides a short-cut that yields equivalent outcomes. Such short-cuts may have been favoured in our evolution because they have lower operating costs, were more easily available to natural selection or are capable of reaching decisions faster. In fact, there are many examples of biases in human judgement and decision-making that seem to be adaptive precisely because they offer simple heuristics that deceive us into fitness-maximizing behaviour[18,20].

The finding that the optimal level of bias increases with the magnitude of uncertainty is especially intriguing. It suggests that we should expect extreme levels of overconfidence (hubris) or underconfidence (fear) precisely when we are dealing with unfamiliar or poorly understood strategic contexts. We predict that where the value of a prize sufficiently exceeds the costs of competing, overconfidence will be particularly prevalent in some very important domains that have inherently high levels of uncertainty, including international relations (where events are complex and distant and involve foreign cultures

and languages), rare or unpredictable phenomena (such as natural disasters and climate change), novel or complex technologies (such as the Internet bubble and modern financial instruments) and new and untested leaders, allies and enemies. Although overconfidence may have been adaptive in our past, and may still be adaptive in some settings today, it seems that we are likely to become overconfident in precisely the most dangerous of situations.

Received 27 May; accepted 25 July 2011.

1. Kanter, R. M. *Confidence: How Winning Streaks and Losing Streaks Begin and End* (Crown Business, 2004).
2. Taylor, S. E. & Brown, J. D. Positive illusions and well-being revisited: separating fact from fiction. *Psychol. Bull.* **116,** 21–27 (1994).
3. Taylor, S. E. *Positive Illusions: Creative Self-Deception and the Healthy Mind* (Basic Books, 1989).
4. Peterson, C. *A Primer in Positive Psychology* (Oxford Univ. Press, 2006).
5. Wrangham, R. W. Is military incompetence adaptive? *Evol. Hum. Behav.* **20,** 3–17 (1999).
6. Trivers, R. L. The elements of a scientific theory of self-deception. *Ann. NY Acad. Sci.* **907,** 114–131 (2000).
7. Trivers, R. *Deceit and Self-Deception: Fooling Yourself the Better to Fool Others* (Allen Lane, 2011).
8. McKay, R. T. & Dennett, D. C. The evolution of misbelief. *Behav. Brain Sci.* **32,** 493–510 (2009).
9. Tuchman, B. W. *The March of Folly: From Troy to Vietnam* (Alfred A. Knopf, 1984).
10. Camerer, C. & Lovallo, D. Overconfidence and excess entry: an experimental approach. *Am. Econ. Rev.* **89,** 306–318 (1999).
11. Malmendier, U. & Tate, G. CEO overconfidence and corporate investment. *J. Finance* **60,** 2661–2700 (2005).
12. Johnson, D. D. P. *Overconfidence and War: The Havoc and Glory of Positive Illusions* (Harvard University Press, 2004).
13. Johnson, D. D. P. & Tierney, D. R. The Rubicon theory of war: how the path to conflict reaches the point of no return. *Int. Secur.* **36,** 7–40 (2011).
14. Sharot, T. *The Optimism Bias: A Tour of The Irrationally Positive Brain* (Pantheon, 2011).
15. Johnson, D. D. P. & Levin, S. A. The tragedy of cognition: psychological biases and environmental inaction. *Curr. Sci.* **97,** 1593–1603 (2009).
16. Akerlof, G. A. & Shiller, R. J. *Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Global Capitalism* (Princeton Univ. Press, 2009).
17. Fudenberg, D. & Tirole, J. Perfect Bayesian equilibrium and sequential equilibrium. *J. Econ. Theory* **53,** 236–260 (1991).
18. Gigerenzer, G. *Adaptive Thinking: Rationality in the Real World* (Oxford Univ. Press, 2002).
19. Nettle, D. in *Emotion, Evolution and Rationality* (eds Evans, D. & Cruse, P.) 193–208 (Oxford Univ. Press, 2004).
20. Haselton, M. G. & Nettle, D. The paranoid optimist: an integrative evolutionary model of cognitive biases. *Pers. Soc. Psychol. Rev.* **10,** 47–66 (2006).
21. Cosmides, L. & Tooby, J. Better than rational: evolutionary psychology and the invisible hand. *Am. Econ. Rev.* **84,** 327–332 (1994).
22. Fama, E. F. & Roll, R. Some properties of symmetric stable distributions. *J. Am. Stat. Assoc.* **63,** 817–836 (1968).
23. Fearon, J. D. Rationalist explanations for war. *Int. Organ.* **49,** 379–414 (1995).
24. Maynard Smith, J. & Parker, G. The logic of asymmetric contests. *Anim. Behav.* **24,** 159–175 (1976).
25. Parker, G. A. Assessment strategy and the evolution of fighting behaviour. *J. Theor. Biol.* **47,** 223–243 (1974).
26. Enquist, M. & Leimar, O. Evolution of fighting behaviour: decision rules and assessment of relative strength. *J. Theor. Biol.* **102,** 387–410 (1983).
27. Nowak, M. A. *Evolutionary Dynamics: Exploring the Equations of Life* (Belknap Press, 2006).
28. Enquist, M. & Leimar, O. The evolution of fatal fighting. *Anim. Behav.* **39,** 1–9 (1990).
29. Johnson, D. D. P., Weidmann, N. B. & Cederman, L.-E. Fortune favours the bold: an agent-based model reveals adaptive advantages of overconfidence in war. *PLoS ONE* **6,** e20851 (2011).
30. McDermott, R., Fowler, J. H. & Smirnov, O. On the evolutionary origin of prospect theory preferences. *J. Polit.* **70,** 335–350 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Barrett, D. Blumstein, L.-E. Cederman, D. Fessler, P. Gočev, M. Haselton, D. Nettle, J. Orbell, K. Panchanathan, M. Price, D. Tierney, R. Trivers, N. Weidmann and R. Wrangham for discussions and help leading to this paper.

Author Contributions D.J. and J.F. conceived the study. J.F. performed the modelling. D.J. and J.F. analysed the results, revised the models and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.J. (dominic.johnson@ed.ac.uk).