Causes of Confidence in Conflict

Robin Hanson*
Department of Economics
George Mason University†

June 2006

Abstract

In a simple model of conflict, two agents fight over a fixed prize, and how hard they fight depends on what they believe about their abilities. To this model I add "pre-agents," representing parents, leaders, or natural selection, who choose each agent's confidence in his ability. Depending on the reason for such confidence, I find five different patterns in how confidence varies with ability. Agents who estimate their ability with error have under-confidence when ability is high and over-confidence when ability is low, while strategic commitment incentives induce the opposite pattern. Agents who misjudge their value for the prize, relative to their cost of effort, induce an over- or under-confidence that is independent of ability, while cooperating pre-agents choose extreme under-confidence. Agents who use confidence to signal ability have a relatively uniform over-confidence.

Introduction

Overconfidence is a important feature of human behavior, and conflict is an important context of human behavior. Furthermore, overconfidence is considered to be especially relevant on the context of conflict (Johnson, 2004).

There is a vast literature detailing the existence and implications of overconfidence, a large literature identifying proximate psychological causes of overconfidence, but only a small literature considering possible functional reasons for overconfidence. Possible functions of overconfidence include as a way to commit to future behavior (Hvide, 2002; Kyle & Wang, 1997), as a way to convince others of one's ability (Trivers, 2000), and as a way to correct for value errors (Haselton & Funder, 2006).

*For comments I thank ?? I thank the Center for Study of Public Choice, and the Mercatus Center for financial support.

†*rhanson@gmu.edu* http://hanson.gmu.edu 703-993-2326 FAX: 703-993-2323 MSN 1D3, Carow Hall, Fairfax VA 22030

This paper explores several possible functions of overconfidence, in a simple previously-elaborated model of conflict (Hirshleifer, 1995). In the basic model, two agents of varying ability expend effort to gain a larger share of fixed pie. This model is extended by introducing the possibility of commitment to over- or under-confidence in ability.

Each agent in the conflict is matched with a "pre-agent," who shares some but perhaps not all of his agent's preferences. Before the conflict, each pre-agent, knowing the agents' true abilities, chooses how confident his agent will be about his ability. The other agent also learns about this confidence before the conflict. "Pre-agents" might represent parents instilling beliefs in their children, or natural selection choosing belief tendencies.

Looking at five different reasons for over- or under-confidence, I find five different patterns relating agent confidence and ability. First, it is clear that if agents simply had error-prone but rational estimates of their abilities, they would tend to over-confidence when they had low ability, and would tend to under-confidence when they had high ability (Klayman, Soll, & Gonzalez-Vallejo, 1999). Second, I find that even when pre-agent and agent preferences are identical, agents gain by committing to over-confidence when dealing with less-able agents, and to under-confidence when dealing with more-able agents.

Third, I find that if an agent misjudges the value of the prize to be won, relative to the cost of conflict, his pre-agent can correct for this error via an ability-independent level of over- or under-confidence. Fourth, I find that pre-agents who cooperate to maximize their joint payoff choose the lowest feasible confidence for all ability levels, in order to minimize the cost of conflict. Fifth, I find that pre-agents who also care about convincing an audience that the agent is of high ability would choose over-confidence for all agents except those of the absolute lowest ability.

Finally, I briefly put these results in the context of what we know about real human confidence.

Basic Conflict

Consider a simple conflict between two agents over a unit pie. This could represent people fighting over a parking spot, a hunted prey, a mate, or anything else of value.

In the basic conflict game, two agents, 1 and 2, first gain common knowledge of their abilities $a_1, a_2 \in [\underline{a}, \bar{a}] \subset R^+$. Second, agent 1 chooses strength $S_1 \in R^+$ while agent 2 simultaneously chooses strength $S_2 \in R^+$. Finally, agent 1 wins fraction

$$F_1 = \frac{S_1}{S_1 + S_2}$$

of the pie, while agent 2 wins the other fraction $F_2 = 1 - F_1$. (Alternatively, fraction F_i could describe agent i's chance of winning the entire pie.) Each agent also suffers a cost of strength

$$C_i = \left(\frac{S_i}{a_i}\right)^{\alpha},$$

which depends on his ability a_i. Costs are convex, so $\alpha \geq 1$, and agent utility is simply

$$U_i = F_i - C_i,$$

pie fraction minus cost of strength.

Given common knowledge of agent abilities a_i, there is a unique pure strategy Nash equilibrium here, and there are no mixed strategy equilibria. In the unique equilibrium, agent 1's strategic reaction function is given by

$$S_1^{\alpha-1} = \frac{a_1^\alpha S_2}{\alpha(S_1+S_2)^2},$$

and a symmetric equation holds for S_2. The strategic dependence dS_1/dS_2 is positive when $F_1 > 1/2$ (and $a_1 > a_2$), and negative when $F_1 < 1/2$ (and $a_1 < a_2$). In this equilibrium,

$$S_1^\alpha = \frac{a_1^{1+\alpha} a_2}{\alpha(a_1+a_2)^2},$$

$$F_1 = \frac{a_1}{a_1+a_2},$$

symmetric expressions hold for agent 2, and $C_1 = C_2$.

Choosing Confidence

Imagine that before a basic conflict between two agents, some other pair of agents, call them pre-agents, choose the beliefs b_i of the basic agents about their abilities a_i. Pre-agents could represent parents instilling beliefs in children, military leaders instilling beliefs in their troops, or natural selection encoding tendencies toward over or under confidence.

At this confidence choice stage, pre-agent 1, knowing both abilities a_i, chooses a belief $b_1 \in R^+$ which agent 1 will simply accept as his value of a_1. After the beliefs b_i are chosen, basic agents gain common knowledge of those beliefs, and then they have a basic conflict.

When pre-agent 1 chooses $b_1 > a_1$ we will say that he chooses *over-confidence* for agent 1, while when he chooses $b_1 < a_1$ we will say that he chooses *under-confidence*. Note that agent 1 does *not* have a rational confidence derived from knowing pre-agent behavior; he instead acts as if does not anticipate deceptive pre-agents. Agent 2 acts symmetrically.

Let us first consider the case where each pre-agent shares the preferences of his respective agent, except that he may place a different weight w on the importance of winning the pie, relative to the cost of strength. Substituting the equilibrium strategies for basic conflict, we find the utility of pre-agent 1 to be

$$\tilde{U}_1 = wF_1 - C_1 = \frac{wb_1}{b_1+b_2} - \frac{b_1^{1+\alpha} b_2}{\alpha a_1^\alpha (b_1+b_2)^2}.$$

A symmetric expression holds for $\tilde{U}_2$.

In this confidence-choosing stage there are also no mixed strategy equilibria and the pure strategy equilibrium is unique. Equilibrium behavior satisfies $b_1 = \tilde{b}(a_1, a_2)$, $b_2 = \tilde{b}(a_2, a_1)$, and

$$w \left(\frac{a_1}{b_1}\right)^{\alpha} = 1 + \frac{b_2 - b_1}{\alpha(b_1 + b_2)}.$$

Thus when pre-agent and agent preferences are identical, with $w = 1$, then each basic agent is over-confident, with $b_1 > a_1$ when he is more able, where $a_1 > a_2$ and $b_1 > b_2$. He is under-confident, with $b_1 < a_1$, when he is less able, where $a_1 < a_2$ and $b_1 < b_2$.

This makes sense since the strategic dependence dS_1/dS_2 in the basic conflict game says that committing to raising one's strength induces the other agent to reduce his strength, but only when that other agent's strength is smaller. When the other agent is instead more able it pays to commit to a smaller strength.

The effect of non-unit weight w and ability a_1 is behaviourly equivalent to the effect of a unit weight $w' = 1$ and ability $a'_1 = w^{1/\alpha} a_1$. Thus a weight above one induces a uniform over-confidence, giving the same effect over the whole range of ability. A weight below one induces a uniform under-confidence.

If these pre-agents could cooperate with each other to maximize their joint payoff, what confidence would they choose? Their joint payoff is

$$\tilde{U}_1 + \tilde{U}_2 = w - C_1 - C_2,$$

which is clearly maximized by minimizing the costs of strength C_1 and C_2. Since agents with low ability choose low strength, then extreme under-confidence, as in $b_1, b_2 \to 0$, would completely eliminate the costs of strength. Thus cooperating pre-agents would choose confidence levels as low as feasible.

Signaling Via Confidence

Let us now consider pre-agents who care not just about winning the pie and losing costs of strength, but who also care about using confidence to impress an audience about agent ability. This audience might, for example, consist of potential mates or allies.

That is, there exists an audience that is initially ignorant of the abilities value a_1, a_2, other than knowing they lie in $[\underline{a}, \bar{a}] \subset R^+$. This audience is, however, aware of the pre-agents and their behavior, and knows that in equilibrium confidence depends on ability via $b_1 = \hat{b}(a_1, a_2)$ and $b_2 = \hat{b}(a_2, a_1)$. After observing the confidence values b_1, b_2, the audience can invert this dependence to infer ability from confidence, via $a_1 = \hat{a}(b_1, b_2)$ and $a_2 = \hat{a}(b_2, b_1)$, where $b_1 = \hat{b}(\hat{a}(b_1, b_2), \hat{a}(b_2, b_1))$.

If pre-agent 1 cares an amount k about impressing this audience, then he has utility

$$\tilde{U}_1 = wF_1 - C_1 + k\,\hat{a}(b_1, b_2).$$

$\tilde{U}_2$ has a symmetric form. When $k = 0$ this reduces to the previous case, so $\hat{b}(a_1, a_2) = \tilde{b}(a_1, a_2)$. And even when $k > 0$, the behavior of the worst types reduces to the previous case, with $\hat{b}(\underline{a}, a_2) = \tilde{b}(\underline{a}, a_2)$. Since his type will be revealed in a separating equilibrium, the worst type ignores signaling incentives.

Better types $a_1 > \underline{a}$, however, do distort their choices in order to distinguish themselves from worst types, inducing an $\hat{a}(b_1, b_2)$ satisfying

$$k \frac{b_1}{\hat{a}} \frac{\partial \hat{a}}{\partial b_1} = \frac{b_1 b_2}{(b_1 + b_2)^2} \left(-1 + \frac{b_1^\alpha}{w\, a_1^\alpha} \left(1 + \frac{b_2 - b_1}{\alpha(b_1 + b_2)} \right) \right).$$

Since for $k > 0$ higher confidence should signal higher ability, the left-hand side of this equation is positive. Thus the value of the ratio b_1/a_1 in the right-hand expression will be increased relative to the $k = 0$ case.

In summary, giving pre-agents an additional incentive to have their agents appear to be of high ability induces them to choose higher confidence. This effect is less for agents of low ability, and more for agents of high ability.

Discussion

In the context of a simple model of conflict, I have introduced a choice of confidence, which can result in over-confidence, under-confidence, or accurate confidence. We have also considered four possible mechanisms which can induce inaccurate confidence. To these four we can trivially add a fifth mechanism, simple error. Let us review how these different mechanisms would make confidence depend on ability.

First, consider simple error. If agents had a noisy signal of their own ability, then rational high ability agents would tend to have under-confidence in their ability, while low ability agents would tend to have over-confidence. Second, there is a strategic effect, whereby confidence variations can let agents commit to strength in a conflict. This strategic effect induces over-confidence in high ability agents, and under-confidence in low ability agents, the opposite pattern from simple error.

Third, if agents have mistakes in their preferences, either neglecting or incorrectly weighing some costs or benefits of conflict, such mistakes can be corrected for by a uniform level of over- or under-confidence in ability. Fourth, when the two sides could cooperate to maximize their joint payoff, they would choose extreme under-confidence at all ability levels. Fifth, an incentive to gain from signaling higher ability would induce over-confidence at all but the lowest ability levels, and more over-confidence at higher ability levels.

These differing reasons for inaccurate confidence are simple and distinct enough to tempt us to try to compare them with our data on over- and under-confidence. While there seems to be no specific data yet on confidence relative to ability in the context of conflict, in other contexts overconfidence seems to be the general rule, tempered often by under-confidence among those of highest ability (Klayman et al., 1999).

If this empirical pattern were to also hold for confidence in conflict, our strongest conclusion would probably be that errors in ability estimation make an important contribution.

In addition, if the simple model of conflict explored in this paper were a reasonable guide, we would guess that the two factors most likely to contribute substantially to the remaining pattern are signaling incentives and over-estimates of the value of the pie, relative to the cost of conflict. Only these mechanisms can induce an over-confidence that is relatively uniform across the ability spectrum. The mechanisms of commitment or cooperation do not produce similar enough patterns to be plausible here.

Without a plausible explanation for a systematic tendency to underestimate the cost of conflict, all things considered the signaling theory seems the most reasonable explanation for confidence in conflict. Of course even if so, there remains the question of why ordinary conscious analysis would tend to neglect the signaling value of conflict, so that overconfidence is needed to correct for such errors. And there remains the question of why we cannot notice and correct for our errors in confidence.

Conclusion

When two agents fight over shares of a fixed pie, how hard they fight depends on what they believe about their relative abilities. "Pre-agents" representing those agents might therefore want to adjust the confidence of their agents in their own ability, in order to better cooperate, to gain strategic benefits of commitment, or to correct for errors in the agent values. Agent value errors can either ignore the value of signaling ability to an audience, or incorrectly weigh the cost of conflict relative to the value of the prize to be gained. A simple model of confidence during conflict, when compared to the inadequate data available, suggests that, in addition to errors in estimating ability, agent ability signaling, together with an unexplained tendency to undervalue such signaling, is the most likely explanation for typical patterns of overconfidence.

References

Haselton, M. G., & Funder, D. C. (2006). The Evolution of Accuracy and Bias in Social Judgment. In Schaller, M., Simpson, J. A., & Kenrick, D. T. (Eds.), *Evolution and Social Psychology*. Psychology Press, New York.

Hirshleifer, J. (1995). Anarchy and its Breakdown. *Journal of Political Economy*, *103*(1), 26–52.

Hvide, H. K. (2002). Pragmatic Beliefs and Overconfidence. *Journal of Economic Behavior and Organization*, *48*(1), 15–28.

Johnson, D. D. (2004). *Overconfidence and War*. Harvard University Press, Cambridge, Massachusetts.

Klayman, J., Soll, J. B., & Gonzalez-Vallejo, C. (1999). Overconfidence: It Depends on How, What, and Whom You Ask. *Organizational Behavior and Human Decision Processes*, *79*(3), 216247.

Kyle, A. S., & Wang, F. A. (1997). Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?. *Journal of Finance*, *52*(5), 2073–2090. overconfidence acts like a commitment device in a standard Cournot duopoly.

Trivers, R. (2000). The Elements of a Scientific Theory of Self-Deception. In LeCroy, D., & Moller, P. (Eds.), *Evolutionary Perspectives on Human Reproductive Behavior*, Vol. 907 of *Annals of the New York Academy of Sciences*.