

Jonathan Birch

How cooperation became the norm

**Article (Accepted version)
(Refereed)**

Original citation:

Birch, Jonathan (2014) How cooperation became the norm. [Biology and Philosophy](#), 29 (3). pp. 433-444. ISSN 0169-3867 (In Press)]

DOI: [10.1007/s10539-013-9409-8](https://doi.org/10.1007/s10539-013-9409-8)

© 2014 [Springer Netherlands](#)

This version available at: <http://eprints.lse.ac.uk/61881/>

Available in LSE Research Online: May 2015

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

How cooperation became the norm

Kim Sterelny, Richard Joyce, Brett Calcott, and Ben Fraser (eds): *Cooperation and Its Evolution*. MIT Press, Cambridge MA, 2013, 608 pp, \$55, ISBN: 978-0-262-01853-1

Jonathan Birch¹

(forthcoming in *Biology and Philosophy*; the final publication is available at link.springer.com)

Abstract

Most of the contributions to *Cooperation and Its Evolution* grapple with the distinctive challenges presented by the project of explaining human sociality. Many of these puzzles have a ‘chicken and egg’ character: our virtually unparalleled capacity for large-scale cooperation is the product of psychological, behavioural, and demographic changes in our recent evolutionary history, and these changes are linked by complex patterns of reciprocal dependence. There is much we do not yet understand about the timing of these changes, and about the order in which different aspects of human social psychology (co-)evolved. In this review essay, I discuss four such puzzles the volume raises. These concern punishment and norm-psychology, moral judgement and the moral emotions, hierarchy and top-down coercion, and property rights and legal systems.

Key words

Human evolution, cooperation, norms, punishment, morality, emotions, hierarchy, property

Humans are part of the natural world and a product of natural processes. Yet we are a strange biological phenomenon, set apart from our fellow great apes by the richness of our social,

¹ Christ’s College, St Andrew’s Street, Cambridge, CB2 3BU, UK.

Email: jgb37@cam.ac.uk

cognitive, and ethical lives. A central task for naturalistic philosophy is to explain how the processes of biological and cultural evolution have led to the psychological and behavioural capacities that make humans unique. *Cooperation and Its Evolution* makes real progress on this task. Across 26 dizzyingly eclectic but consistently interesting chapters, the volume brings together a wealth of expertise from philosophy, anthropology, economics, psychology, and behavioural ecology in pursuit of integrative explanations that no single discipline can hope to provide alone.

Not all the contributions are concerned with humans. Some chapters focus on sociality in non-human animals, including eusocial insects, birds, and bacteria. But although specialists in non-human behavioural ecology may find some valuable material here, a brief glance at the table of contents is enough to see that *Homo sapiens* dominates this collection. Accordingly (and with apologies to the authors of the chapters I won't discuss) this review essay will also be anthropocentric in its focus. I will highlight four key explanatory puzzles concerning human social evolution that the papers in this volume raise—puzzles that appear quite different at first glance, but that in fact have a shared underlying structure.

Chickens and eggs

Modern evolutionary biologists lose little sleep over the traditional 'chicken and egg' problem. Once we recognize that speciation happens gradually, we see that the question is biologically uninteresting: at time t_1 , there were unambiguous non-chickens developing from non-chicken eggs; at some later time t_2 , following a speciation event, there were unambiguous chickens developing from chicken eggs.² In between was a region of borderline cases, and it is futile to seek the definitive first chicken, or the definitive first chicken egg, at some determinate point in this region. Theorists of vagueness may still worry³, but the appearance of a substantive biological problem dissolves.

Nevertheless, problems of a broadly 'chicken and egg' character continue to trouble evolutionary biologists, for the natural world is full of relationships of intricate mutual

² The chicken is technically a subspecies of the red jungle fowl (*Gallus gallus*). Readers of a pedantic disposition are welcome to substitute '*G. gallus*' for 'chicken' here and elsewhere in the essay.

³ See, for example, the curious debate in the pages of *Mind* between Teichmann (1991), Sorensen (1992) and Waller (1998). The debate turns out to hinge on the meaning of 'egg'.

dependence that cry out for explanation. Perhaps most famously, many flowering plants are dependent on specific pollinators, which are in turn adapted to pollinate that specific plant. Neither can prosper without the other, and yet there must have been a time at which neither existed. How did evolution get from there to here? Which species, if either, evolved first? Such problems are not rendered trivial by the bare fact of gradual speciation. Even so, Darwinian gradualism provides the basic recipe for a solution. If *Xs* evolved gradually from proto-*Xs*, and proto-*Xs* evolved gradually from proto-proto-*Xs*, and if proto-proto-*Xs* were *not* dependent on pre-existing proto-proto-*Ys*, then we can start to see how *Xs* and *Ys* might have *co-evolved*, with each incremental change in one enabling a stepwise change in the other. Of course, this provides only the bare bones of an explanation. In each case, we still need to explain how and why this co-evolutionary ratcheting occurred, and this is where the real explanatory work gets done.

Human social evolution presents a host of highly non-trivial ‘chicken and egg’ problems. The story of our divergence from the other great apes over the past two million years is the story of an explosion in the scale and sophistication of cooperation, accompanied by rapid and radical changes in our cognitive architecture, life-history characteristics, diet, demography, and ecological environment. It seems likely that all of these changes were interrelated, and this leads naturally to ‘Which came first?’-type questions. As with plants and their pollinators, the quick answer is that all these changes probably occurred gradually and concurrently by means of co-evolutionary processes. But that, obviously, is too quick. For we still want to know how and why these co-evolutionary processes took place. We want to know which changes enabled which others, what the incremental steps were, and what evolutionary processes drove the transition. We also want to know why no other lineage—in the great apes or anywhere else—has undergone such an extraordinary transformation.

Many of the articles in *Cooperation and Its Evolution* grapple with these difficult issues. In particular, four ‘chicken and egg’ problems stand out. These concern punishment and norm-psychology, moral judgement and the moral emotions, hierarchy and top-down coercion, and property rights and legal systems. In some of these cases, the author sees the problem and suggests a possible solution. In others, the problem lurks in the background, unnoticed by the author. In all four cases, there is much we do not yet understand about how (co-)evolution actually solved the problem.

Punishment and norm-psychology

One might initially find it surprising that, in a book ostensibly about cooperation, one of the longest entries in the index is for ‘Punishment’. This reflects a growing theoretical consensus that punishment was critical in stabilizing cooperation between non-relatives in early human populations—a point made forcefully here in the chapters by Fiery Cushman and by Maciek Chudek and colleagues. It would be both implausible and somewhat misanthropic, however, to suggest that modern humans only ever cooperate through fear of punishment. On the contrary (and as Chudek et al. emphasize), the evidence suggests that we have a sophisticated ‘norm-psychology’: a dedicated faculty that acquires prevailing social norms and motivates both our own compliance and our imposition of sanctions on those who fail to comply (Sripada and Stich 2006).

Here, then, is our first ‘chicken and egg’ problem. Plausibly, punishment played a key role in stabilizing cooperative social norms. Once such norms were stabilized, it is straightforward to see why humans might have evolved an internalized norm-psychology, so that the motivational role initially played by the threat of external sanction came to be played by emotional responses to perceived norm violations in oneself and others. Yet, in modern humans, punishment is itself motivated in large part by emotional responses generated by our evolved norm-psychology. So which came first?

One possible solution is that our basic capacity for punishment pre-dates the motivational architecture that now underpins it. The tenability of this response will depend on what we take a ‘basic capacity for punishment’ to consist in. In modern societies, we recognize a distinction between punishment and mere retaliation. If you pull my hair and I pull yours, I am merely retaliating—I am not punishing you. Punishment proper is entangled with the concept of justice; it involves a judgement on the part of the punisher that a socially accepted norm has been violated and that the violation warrants punishment. On this conception, then, punishment requires a cognitive capacity to recognize norm violations *qua* norm violations and a motivational structure that impels us to respond to them *because they are norm violations*—and not just, for example, because retaliation is in our prudential interests. Given this, it would appear that even a basic capacity for punishment implicates several aspects of the norm-psychology it is frequently invoked to explain.

Yet co-evolution may come to the rescue, provided we are willing to postulate a gradual transition from mere retaliation to full-blown punishment, complete with its distinctive cognitive elements. The story, in very broad terms, would have had three stages.⁴ We began with nothing more than violent retaliation, triggered by certain behaviours that deviated from established behavioural regularities. This retaliation initially evolved because it was in the evolutionary interests of the retaliator: it asserted their dominance, weakened their rivals, and deterred future attempts to steal their resources. Such behaviours are well documented in chimpanzees (de Waal 1982). This systematic, violent retaliation can be viewed as a form of niche construction, since it alters the selective regime agents subsequently face. In particular, if violent retaliation is a major feature of social life, agents who are able to forestall retaliation by predicting the behaviours that trigger it will be at an advantage. Consequently, in the second stage, natural selection built a dedicated system for acquiring the social norms one needs to follow in order to avoid retaliation and for motivating behaviour that complies with these norms; in other words, selection assembled our norm-psychology. But this was not the end of the story. For norm-psychology brought with it the capacity to grasp and accept norms (or ‘meta-norms’) regarding retaliation itself. When our accepted retaliatory norms mandated a violent response to certain behaviours, we came to conceptualize these behaviours as *deserving sanction*, and retaliation gradually gave way to full-blown punishment. If this story is on the right lines, it is a nice evolutionary irony: a psychological faculty that evolved in response to precursors of punishment was itself the key enabler for the evolution of punishment as we now conceive it.

One remaining question is why this transition only occurred in hominins, given that the basic platform—i.e., a capacity for systematic, violent retaliation—is clearly present in other great apes. Social learning may provide the crucial piece of the puzzle. For, as Cushman persuasively argues, human norm-psychology represents an impressive cognitive achievement: to learn the norms of one’s local community—to learn, in particular, what kind of behaviour evokes what degree of retaliation—is no small epistemic feat. It demands significant pre-existing capacities for social learning and cultural transmission, and these are capacities that the other great apes do not have to anything like the same degree. Plausibly, then, the evolution of norm-psychology has substantial cognitive pre-requisites, and the other great apes never experienced the cooperative explosion norm-psychology facilitates because they never satisfied these preconditions. This leaves us with the problem of explaining why

⁴ Kitcher (2011, Ch 2) tells a story not too dissimilar to this.

hominins evolved their extraordinary capacity for social learning in the first place (cf. Sterelny 2012). But as long as our preferred explanation for this capacity does not invoke a pre-existing norm-psychology, the threat of circularity diminishes.

Moral judgement and the moral emotions

It is uncontroversial that moral judgement in humans is frequently accompanied by moral emotions such as guilt, shame, indignation, outrage, and resentment. More controversially, a substantial body of evidence suggests that in many cases ‘the emotional tail wags the rational dog’: explicit moral deliberation often does little more than provide *post hoc* rationalizations for snap judgements made in the emotional heat of the moment (see the chapters by Kelly and Tangney et al. in this volume; see also Haidt 2001). Note, however, that we can accept an intimate connection between moral emotions and moral judgement and yet still disagree about the *direction of psychological explanation* between them; and indeed one finds two opposing positions on this question in the contemporary literature.

On the one hand, there are the ‘emotionists’ (e.g. Nichols 2004; Prinz 2007), who maintain that emotions are psychologically prior to moral judgements—i.e., that the phenomenon of moral judgement is to be explained (either causally or constitutively⁵) in terms of characteristic emotional responses to norm violations. They further hypothesize these emotional responses are constructed in cognitive development from fundamentally non-moral emotional ingredients.⁶ On the other hand, there is the broadly cognitivist view (defended in this volume and elsewhere by Richard Joyce) that moral emotions are *not* reducible to mere complexes of non-moral emotions; that they contain distinctive cognitive elements; and that, as a consequence, they can only be experienced by agents with pre-existing competence in the application of moral concepts. On this view, the direction of explanation is reversed: moral judgements, conceived as a largely cognitive phenomenon, are either causally or constitutively implicated in the production of moral emotions.

⁵ This is a point of disagreement between Nichols (who sees the relationship between moral emotions and moral judgement as one of causal relevance) and Prinz (who sees the relationship as one of constitutive relevance). See Prinz 2007 and Prinz and Nichols 2010 for discussion.

⁶ Prinz, for example, characterizes guilt as a variety of sadness—a claim Joyce explicitly criticizes in his contribution.

As both camps are well aware, this debate holds significant implications regarding the *evolution* of moral judgement. For as Jesse Prinz (2009) has noted, the emotionist view dovetails with a by-product hypothesis, on which a capacity for full-blown moral judgement came along more or less for free once we had evolved (i) a (not distinctively moral) capacity for norm acquisition and (ii) a capacity to feel basic, non-moral emotions in response to perceived norm violations. The cognitivist view, by contrast, implies that moral judgement is *not* readily explained as a mere by-product of other cognitive adaptations. For if moral judgement involves the application of distinctively moral concepts, and if these concepts are more than mere projections of fundamentally non-moral emotions, then it seems likely that a substantial amount of further adaptive evolution was required for hominins equipped with non-moral emotions and a basic capacity for norm acquisition to evolve the additional conceptual apparatus that moral judgement requires. The cognitivist view therefore lends itself to adaptationist hypotheses regarding the evolution of morality.

At present, the debate (on both its psychological and evolutionary fronts) appears to have reached a state of deadlock. In his valuable contribution to this volume, Joyce attempts to plot a way forward. The key move, he urges, is to draw a distinction between more and less cognitively demanding conceptions of moral judgement. The less demanding variety—let’s call it moral judgement-lite—is such that it ‘can be built largely from emotional resources’ (p 565); it is, in other words, moral judgement *sensu* Nichols and Prinz. The more demanding variety—let’s call it moral judgement-deluxe—is such that it captures ‘crucial cognitive elements ... like *desert, transgression, practical authority* (etc.)’ (p 566) that emotional resources alone cannot explain. The apparent upshot is that we can have our cake and eat it, with regard to both psychological and evolutionary explanation. We can side with the emotionists regarding moral judgement-lite: we can agree that *this* capacity is constructed in development from purely emotional ingredients and, as such, may well have evolved as a by-product of other cognitive adaptations. But we can also side with the cognitivists regarding moral judgement-deluxe: we can agree that *this* capacity is *not* constructed in development from purely emotional ingredients and therefore calls for a distinctive psychological and evolutionary explanation.

I suspect that card-carrying emotionists are unlikely to welcome this olive branch with open arms. For it is surely a core commitment of their view that it does not merely characterize a ‘less demanding’ variety of moral judgement, stripped of its distinctive cognitive elements, but rather characterizes *all there is* to moral judgement in human beings.

If there turn out to be important elements of human moral psychology that the emotionist account leaves out, then the emotionists have lost the debate. Accepting Joyce's purportedly conciliatory resolution would thus be equivalent to admitting defeat.

Yet although Joyce's distinction between varieties of moral judgement may not be as 'pluralistic' as he makes it out to be, I think it remains a helpful move—and particularly so when our focus turns to diachronic questions. For if we want to explain the evolution of our capacity for moral judgement, an acknowledgement that moral judgement comes in grades, and that these grades vary in their cognitive requirements, seems virtually unavoidable. This leads me to speculate that Joyce's less- and more-demanding sharpenings of the notion of 'moral judgement' might usefully be reinterpreted in diachronic terms, as a conjectural sequence of stages through which the evolutionary transition from basic emotional capacities to full-blown moral judgement proceeded. For example, building on our discussion of norm-psychology in the previous section, we might hypothesize the following eight-stage transition, in which each step involves the addition of a new capacity relevant to moral judgement:

1. The capacity to feel basic emotions (e.g. *anger, sadness*) in response to the perceived actions of oneself and others.
2. The capacity to acquire and accept norms, and to recognize actions as violating or complying with those norms.
3. The capacity to feel basic emotions in response to recognized instances of the violation of (or compliance with) accepted norms.
4. The capacity to acquire and accept 'meta-norms' regarding appropriate responses to the actions of others, and to recognize the responses of oneself and others as violating or complying with these meta-norms.
5. The capacity to feel basic 'meta-emotions' in response to recognized instances of the violation of (or compliance with) meta-norms.
6. The capacity to identify some norms as *moral* rather than merely conventional.
7. The capacity to grasp distinctively moral *concepts* (e.g. *right, good, just, deserving, merited* and their antonyms), and to apply them to recognized instances of violations of (and compliance with) moral norms.
8. The capacity to feel distinctively moral *emotions* (e.g. *guilt, indignation, shame, resentment*) and meta-emotions in response to recognized instances of the violation of (or compliance with) moral norms and meta-norms.

Naturally, emotionists will maintain that, once capacities (1)-(5) are in place, capacities (6)-(8) come along more or less for free, at least as far as biological evolution is concerned. For (as we have seen) their position is that moral emotions and moral concepts are assembled in development from basic emotional ingredients, in conjunction with a general capacity for norm acquisition that is not specific to the moral domain. The upshot is that, once these baseline capacities have evolved, biological evolution falls out of the picture: a mixture of cultural evolution and individual learning can do the rest. Note, however, that this is compatible with the claim that capacities (6)-(8) evolved gradually, provided they did so by means of cultural rather than genetic evolution. Note also that an emotionist who accepts this gradualist hypothesis still has plenty of unanswered genealogical questions to mull over—questions concerning the order in which the capacities evolved, the timing of each stage in human history, and the nature of the cultural-evolutionary processes that brought about each incremental change.

Those with more cognitivist sympathies, meanwhile, will maintain that the gap between capacities (1)-(5) and capacities (6)-(8) is one that no amount of individual learning or purely cultural evolution could bridge. They will therefore postulate that these latter capacities are the products of a dedicated, domain-specific faculty for moral cognition that evolved by genetic evolution, and they will downplay the role of culture in producing these capacities. Again, however, note that this hypothesis too falls well short of a complete account of the evolution of moral judgement. For there is a world of difference between merely hypothesizing *that* some such capacity evolved by genetic evolution and explaining *how and why* it did so. For instance, did the transition proceed through steps (1) to (8) in the order given above, or does this list get the order wrong? Did the relevant capacities evolve sequentially, or did clusters of related capacities co-evolve simultaneously? Was selection at the level of the individual responsible, or was it selection at the level of the group? These, and many other questions, remain almost entirely open. For emotionists and cognitivists alike, the genealogy of moral judgement remains a story we have barely begun to tell.

Hierarchy and top-down coercion

It is widely thought that Pleistocene hunter-gatherer bands were largely egalitarian, in the sense that they did not tolerate systematic differences in access to key economic resources

(Knauft 1991; Boehm 1999). Evidently, something changed. Modern societies are rampantly inegalitarian: a hugely disproportionate share of wealth and resources is concentrated in the hands of a small elite. Depressingly, archaeological evidence in the form of grave goods, dwellings and storage facilities indicates that significant economic inequality has been a widespread feature of human social life since roughly the beginning of the Holocene, 11,700 years ago (Bowles et al. 2010). The masses at the bottom of the pyramid may not like it, but (occasional revolutions aside) the coercive machinery of the state gives them little choice in the matter.

A ‘chicken and egg’ puzzle arises in this context because it appears that systematic economic inequality preceded the first known states by several millennia. Inequality therefore pre-dates the institutional mechanisms that now sustain it. The question this time is not so much ‘Which came first?’ as ‘How can it possibly be that way round?’ Why would the majority in a stateless society accept a disproportionately small share of group resources in the absence of institutionalized coercion? This is the question explored in the chapters by Kim Sterelny and Paul Seabright.

A spectrum of possible answers naturally springs to mind, ranging between extreme pessimism and extreme optimism. At the pessimistic end, there is what we might call (following Coetzee 2007) the ‘Kurosawan Theory’ of the origins of inequality, on which early elites, like the bandits in Kurosawa’s *Seven Samurai*, were nothing more than brutish marauders who extracted tribute from vulnerable and dispersed agrarian communities with threats (often realized) of violence. It is easy enough to imagine how this might have worked in the absence of state-like institutions: if the marauders were mobile and struck without warning, farming communities would have struggled to repel them. It is also easy to see how these proto-tax collectors might have gradually morphed into the elites of recognizable states. At the optimistic extreme, there is the suggestion that elites earned their keep by providing centralized management of large-scale cooperative tasks with a degree of efficiency that an egalitarian society could never equal, so that acquiescence to hierarchy was in the interests of the majority.

In his contribution, Sterelny provides a forceful presentation of this explanatory puzzle, and plots a speculative course between these pessimistic and optimistic extremes. He sides with the Kurosawans in granting that elites are indeed ‘parasites’, in the sense that they exploit the labour of others; but goes on to suggest that they were tolerated by the majority

because the intense and frequent intergroup warfare of the early Holocene put a premium on strong and centralized military leadership. It is (as Sterelny notes) a surprising inversion of the usual group selectionist logic: on this hypothesis, fierce competition at the group level, far from disfavouing free riders, was precisely what enabled them to prosper. This may sound implausible on first hearing—and it would be, if we were imagining two unconditional cultural variants (‘free-ride’ or ‘work’) competing for representation in a population. Free-riders would spread within groups, but groups with more free-riders would be outcompeted by groups with fewer. That, however, is not the hypothesis Sterelny has in mind. Rather, Sterelny is hypothesizing a facultatively expressed social norm (roughly: ‘if born into an elite, demand an unequal share of group resources; if not, acquiesce to that demand’) that, when adopted universally by the members of a particular group, enables the individuals in that group to outperform those in groups with alternative norms.

Although, inevitably, this proposal is largely conjectural, it seems credible at face value. It is not clear to me, however, why Sterelny frames this explanation in group selectionist terms, particularly given his usual preference for individualistic modes of explanation. For it seems to me that his hypothesis can be straightforwardly recast in terms of individual benefit. As I understand it, the thought is that if a chief provides effective military leadership that potentially makes the difference between victory and defeat in battle, then it is profitable (in the long-run) for ordinary labourers to acquiesce to his or her demands for an unequal share of group resources: the expected benefit (i.e. a decreased probability of dying in battle) outweighs the cost. Fundamentally, this is a direct fitness explanation for acquiescence to inequality: norms of acquiescence evolve because, at an individual level, it pays to adopt them. It seems optional, and not necessarily illuminating, to frame the hypothesis in terms of group selection.

Seabright, in what is in effect a chapter-length commentary on Sterelny’s argument, suggests some further alternative hypotheses. Several are variants on the pessimistic theory, highlighting various ways in which systematic exploitation remains a possibility even in the absence of institutionalized coercion. Seabright points out, for instance, that inequality may have begun with one group enslaving another, rather than with stratification within a single group. He also notes that, since large, agrarian societies can afford to lose more of their members than small, forager bands, the credible threat of exclusion from the group may have provided an effective means of coercion. This points to an intriguing possible connection between hierarchy and the redundancy afforded by large group size (cf. Sterelny 2012; Birch

2012). But as Seabright acknowledges with refreshing humility, what we have at this stage is an unsolved puzzle, a plethora of possible solutions, and insufficient evidence to choose between them.

Property rights and legal systems

Property rights lie at the heart of our economic interactions with one another. In modern capitalist societies, these rights are enshrined in law and enforced, where necessary, by legal systems. Yet there is at least some evidence that norms of respect for private property (much like inequality and hierarchy) pre-date the institutions that now underwrite them.⁷ Herbert Gintis, in his contribution (a condensed version of his (2007) article), aims to show how respect for private property could have evolved in the absence of institutional backing. By means of a simple game-theoretic model, Gintis shows that, when violent conflict over resources is costly, there are often stable ‘property equilibria’ in the evolutionary dynamics, at which disputes are resolved non-violently in favour of the agent who currently possesses the contested resource.

Gintis’s argument relies heavily on a territorial analogy. In numerous animal species, conflicts over territory are settled by considerations of incumbency: on recognizing that an area is occupied, an intruder will often flee without putting up a fight. The standard explanation for this behaviour is that, when conflict is costly, it is often in the interests of both parties to settle disputes over territory non-violently; and a pattern of default conflict resolution in favour of the incumbent is one evolutionarily stable means of achieving this outcome (Maynard Smith and Parker 1976). Gintis’s proposal is that respect for private property in human societies evolved from this basic capacity for territoriality—and for much the same reason.⁸

⁷ Gintis points to the ‘endowment effect’, the well-documented psychological tendency of experimental subjects to value a good they currently possess more highly than a materially equivalent good they do not possess (Kahneman et al. 1990). This effect has been recorded in young children and even in non-human primates, suggesting it may have deep evolutionary roots (though see Apicella et al. forthcoming). Even granting this, however, I am sceptical that this tendency indicates recognition of and respect for the possessions of others, as opposed to mere reluctance to part with one’s own.

⁸ Stake (2004) and Krier (2009) have also made proposals along these lines.

While the territorial analogy is certainly suggestive, it is worth pointing out one important disanalogy. Gintis's model, like that of Maynard Smith and Parker, takes incumbency (current possession) as exogenous to the dispute: incumbents and intruders contest future possession of the territory, but they do not contest the matter of who currently possesses it. This is arguably reasonable if our target phenomenon is territoriality in animals. For in a territorial dispute, the parties contest *future* incumbency, but the facts about *current* incumbency will often be epistemically transparent: it is often clear to both parties which of them currently controls the territory and which is the intruder.

If, however, the target phenomenon is respect for property in humans, this assumption seems questionable. For in human property disputes, an agent's claim to *having possession* of the contested object will often be itself contested. *A* may be holding the spear at the moment, but *A* stole it from *B*, so does *B* not remain its rightful possessor? In other words, property rights in contemporary human societies rely heavily on a distinction between possession *de facto* and possession *de jure*; and, when the two come apart, it is possession *de jure* that confers property rights. Yet facts about possession *de jure* depend on legal conventions, which in turn depend on legal institutions. This is one reason to think that the legal chicken pre-dates the property rights egg.

Even so, a gradualist move of the sort we encountered in previous sections may help us reconcile this observation with Gintis's account. It seems plausible enough that early human groups would have settled on a standing convention (perhaps internalized through psychological mechanisms such as the endowment effect) to resolve disputes non-violently in favour of the *de facto* possessor of the contested resource. This would have limited violent confrontation, but it would not have yielded a recognizable system of property rights. In particular, the concept of theft would have had no obvious place in such a society, and most behaviour that we would now regard as theft would have gone unpunished. Much later (indeed, probably some millennia after the Pleistocene-Holocene transition), the emergence of legal institutions enabled the stabilization of a far more complex set of property conventions, leading to the concepts of theft and of possession *de jure*. A simple system of conflict resolution based on *de facto* possession was thus gradually displaced by a complex legal system, sustained by top-down coercion, in which lawful and *de facto* possession often part ways. This, of course, is merely the faintest outline of the transition. As in the preceding three cases, all the interesting details are yet to be filled in.

Conclusion

The structural parallels between the four preceding cases are clear. In all four, we find two features of human social life that now seem inextricably entangled, and we are left wondering how one could ever have evolved without the other. The most promising strategy, in all four cases, is to break the target phenomenon into its constituent parts and attempt to construct a gradual transition in which those parts were cumulatively assembled, each successive change in one enabling further alterations in the other. It is an explanatory strategy that has reaped dividends in other areas of biology, and it will no doubt do so in the human case too. But there are so many pieces of the jigsaw we do not yet have—and so many alternative ways to fill in the gaps.

In sum, *Cooperation and Its Evolution* is essential reading for anyone who seeks a better understanding of the puzzles we face in explaining human social evolution and a sharper picture of the space of possible solutions.

References

- Apicella CL, Azevedo EM, Christakis NA, Fowler JH (forthcoming) Evolutionary origins of the endowment effect: evidence from hunter-gatherers. *Am Econ Rev*
- Birch J (2012) Collective action in the fraternal transitions. *Biol Philos* 27:363-380
- Boehm C (1999) *Hierarchy in the forest: the evolution of egalitarian behavior*. Harvard University Press, Cambridge MA
- Bowles S, Smith EA, Mulder MB (2010) The emergence and persistence of inequality in premodern societies. *Curr Anthropol* 51:7-17
- Coetzee JM (2007) *Diary of a bad year*. Harvill Secker, London
- de Waal F (1982) *Chimpanzee politics: power and sex among apes*. Harper and Row, New York
- Gintis H (2007) The evolution of private property. *J Econ Behav Organ* 64:1-16
- Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol Rev* 108:814-834
- Kahneman D, Knetsch JL, Thaler RH (1990) Experimental tests of the endowment effect and the Coase theorem. *J Pol Econ* 98:1325-1348
- Kitcher P (2011) *The ethical project*. Harvard University Press, Cambridge MA
- Knauft BM (1991) Violence and sociality in human evolution. *Curr Anthropol* 32:391-428
- Krier JE (2009) Evolution theory and the origin of property rights. *Cornell L Rev* 95:139-159
- Maynard Smith J, Parker GA (1976) The logic of asymmetric contests. *Anim Behav* 24:159-175
- Nichols S (2004) *Sentimental rules: on the natural foundations of moral judgment*. Oxford University Press, New York
- Prinz JJ (2007) *The emotional construction of morals*. Oxford University Press, Oxford
- Prinz JJ (2009) Against moral nativism. In: Murphy D, Bishop M (eds) *Stich and his critics*. Wiley-Blackwell, Chichester, pp 167-189

- Prinz JJ, Nichols S (2010) Moral emotions. In: Doris J (ed) *The moral psychology handbook*. Oxford University Press, Oxford, pp 111-146
- Sorensen RA (1992) The egg came before the chicken. *Mind* 101:541-542
- Sripada C, Stich S (2006) A framework for the psychology of norms. In: Carruthers P, Laurence S, Stich S (eds) *The innate mind vol. 2: culture and cognition*. Oxford University Press, New York, pp 280-301
- Stake JE (2004) The property 'instinct'. *Phil Trans R Soc Lond B* 359:1763-1774
- Sterelny K (2012) *The evolved apprentice: how evolution made humans unique*. MIT Press, Cambridge MA
- Teichmann R (1991) The chicken and the egg. *Mind* 100:371-372
- Waller D (1998) The chicken and her egg. *Mind* 107:851-853