The evolution of cooperation
A white paper prepared for the John Templeton foundation

Contents

Introduction

You probably take cooperation for granted. You'd be excused for doing so – cooperation is all around us. Children team up to complete a project on time. Neighbours help each other mend fences. Colleagues share ideas and resources. The very fabric of our society is cooperative. We divide up tasks, with farmers producing food, policemen upholding laws, teachers teaching, so that we may all share the benefits of a functioning society without any one person having to master all domains.

What's more, cooperation is logical, at least to you and me. Two hands are better than one, as the saying goes. If you want something another person has, it makes sense that you might share something of your own. Division of labour efficient. If you have a reputation as a cooperative person, others will likely help you down the line. Cooperation is a straightforward way to achieve more than you ever could on your own.

And last but not least, cooperation is simply *good*. Regardless of its utilitarian value, there is something intuitively positive about cooperation, something we should strive to do in the absence of any benefits. Cooperation is everywhere, and this is as it should be.

And yet, despite all of that, to biologists, cooperation makes little sense at all. It is a profound puzzle that dates all the way back to Darwin.

Darwin's theory of design
Few people realize that Darwin's theory of natural selection (Darwin, 1859) is a theory of design. Design is one of the most salient features of life. Organisms *appear as though they were designed*. A cheetah's legs, each tendon, joint and artery working in harmony, are clearly *made for running* -- the perfect components of a running machine. A human eye is *for seeing*. It's not for eating, or for bouncing -- it is a nearly flawless visual receptacle. Fur is for keeping wolves warm in the snow, teeth are for chewing, and a moth's comically long tongue is for getting nectar out of a flower. From the tiny parts inside a cell, to the longest neck of a giraffe, everywhere we look we see components of living things that appear designed *as if for a purpose*.

Darwin's two great achievements were to demonstrate how organisms gain the appearance of being designed for a purpose *without actually being designed*, and to identify that purpose for which they appear designed.

The process, of course, is natural selection. Organisms have variation in their traits (variation). For example, some have longer or shorter necks. That variation occasionally has an impact on the number of offspring individuals have (differential success linked to variation). For example, longer necked individuals have more babies. And finally, that variation is inherited (inheritance of variation). Together, these three ingredients – variation, differential success linked to the variation, and inheritance of the variation – are necessary and sufficient to cause natural selection.

The product of that process, over time, is that traits that cause individuals to have more babies become more common in the population. Over time, more individuals in the population have long necks, until *everyone* has long necks. At this point, we might call a long neck *an adaptation* – which is another word for the traits

that are associated with having more babies over long periods of time, and are therefore selected. Over very long periods of time, nearly all the traits an organism possesses are those traits that cause them to have more babies (adaptations). Organisms eventually appear as though they were *designed* to have more babies.

That, Darwin argued, was the *purpose* for which organisms appear designed: maximizing number of offspring. Another way to say this is that adaptations are *for* having more offspring. Indeed, an adaptation can't ultimately be *for* anything else. In the immediate sense, sharp teeth may be for biting into prey. But ultimately, they are there because being better at biting prey gets you more food which makes you healthier which makes you have more babies. If it didn't make you have more babies, it *couldn't evolve.* Thus, Darwin's theory was not just a theory of how adaptations come about, but what adaptations are *for* (Gardner, 2017).

The puzzle of cooperation

Thus, we arrive at the puzzle of cooperation. Darwin told us that as a result of selection acting on individuals in the population, organisms should appear designed to maximize their number of offspring.

How then, can we explain a bird helping its sister at the nest? If organisms should maximize their own number of offspring, how could a trait evolve that caused a bird to spend time helping *another* individual raise offspring, when this almost certainly takes away from time it could be spending having its own offspring?

Perhaps, you are thinking this isn't too troubling. If, after all, one odd bird chose to help at the nest, we could ignore it, brush it under the rug. But cooperation is ubiquitous in the natural world. Lions collaborate on hunts, flowers share nectar with bees, and even bacteria produce essential resources that their neighbors benefit from.

Moreover, cooperation in nature goes far beyond a helping hand at the nest. Consider, for example, the extreme case of worker ants. These creatures give up reproduction *altogether* to help the colony. There is no view in which this can be explained as an adaptation *for* having more offspring – they don't have any offspring at all!

This kind of extreme cooperation – the giving up of reproduction altogether – isn't just found in some odd insects in the Amazon. Our very bodies – the bodies of all multicellular organisms – are the product of this extreme sacrifice. The vast majority of cells in your body never make it into the next generation – all forty trillion of them sacrifice for the few sperm or egg cells that do.

Thus, Darwin's theory for what the purpose of adaptations is falls short. It's easy to forget that the theory of natural selection is a *predictive* theory – that is, it makes predictions. One kind of prediction it makes is with regards to what kinds of adaptations we should see in nature. The original theory predicts we should find adaptations only *for* having more offspring – and the very cells in our body give the lie to that!

It's worth pausing and sitting with just how strange cooperation is. A thought experiment might help. Imagine a population of cooperative spiders. These spiders each contribute to some shared resource. Let's say, for example, they gather some food, put it in a communal pot, and then everyone partakes in the big feast. Gathering food takes time and energy, but the benefit is that there is always a big pot of food for everyone. They live this way happily for many generations.

But now, imagine, one day, a mutant spider is born, which doesn't contribute to the pot, never gathers food, and yet still gorges on the communal pot. In other

words, this spider incurs none of the *costs* of the hard work, but reaps all the benefits from others' hard work. We might call this spider *selfish.* This spider will inevitably have more energy, incur fewer risks, etc., and have more offspring than its cooperative compatriots. These offspring, inheriting their mother's selfishness-gene, will, too, be selfish.

In the subsequent generation, then, there will be *more* selfish spiders than before. Again, they will do better, have more babies, and on, until *all the spiders* are selfish, and no one is gathering food. This spread of the selfish mutant is *inevitable*, because the selfish one *always* reaps the benefits of cooperation without incurring any of the costs.

It doesn't matter that you and I can see, from our bird's eye-view, that it would be *wiser* for all the spiders to agree to cooperate. It would be *wiser* for selection to favour the more prudent, cooperative spider, because in the *long run* this will be better for everyone. The spiders have no foresight, and nor does natural selection. It selects whichever trait causes more babies in that generation. Selfishness always wins.

That then, is the puzzle of cooperation. It kept Darwin up at night, and remains a thriving and ever-growing field of research. Why would an organism *ever* cooperate? How could cooperation *ever* remain stable in the face of invading selfish mutants, or cheaters? Why do some organisms cooperate and others act selfishly? What conditions determine when cooperation does and doesn't evolve? Why do humans cooperate? Are we more or less cooperative than other organisms?

This review
That's what this review is about. It covers some of the cutting-edge research on the evolution of cooperation. Niko Tinbergen, one of the few evolutionary biologists to win a Nobel prize, argued that there are four types of questions in biology (Figure 1).

	Dynamic	**Static**
Proximate	Ontogeny (Development)	Mechanism (Causation)
Ultimate	Phylogeny (Evolution)	Function (Adaptation)

Figure 1. Tinbergen's four questions. According to Tinbergen, questions in biology can be divided by whether they are dynamic (how things change over time) or static (how things are) and whether they are proximate (the organism without regard to its past) or ultimate (evolutionary questions).

These four questions map roughly onto Aristotle's four causes.

This review won't focus much on the time-related questions of cooperation – how cooperation develops over time within an individual (ontogeny), and when cooperation has evolved in the history of life (phylogeny).

As is often the case in evolutionary biology, the vast majority of our focus will be the *why*. That's because that's usually the most interesting question – in part, because evolutionary biology is unique in the physical sciences for being able to ask the why question at all! Natural selection creates *apparent* purpose, which means we can ask what things are *for* and *why* they evolved.

At the end, we will turn to the *how*, to dive a little deeper into how cooperation works in the human brain – what mechanisms drive and control it.

The review is roughly split in half, taxonomically. In the first part we consider cooperation *across the tree of life*. This covers research on cooperation *in general*, and also cooperation in non-human organisms. In the second part, we turn to cooperation in humans.

It's often useful to keep human explanations and non-human explanations separate in evolutionary biology. That's not because humans aren't products of evolution, or because the rules of natural selection don't apply to us. It's simply because in *addition* to standard explanations for traits in humans, we are affected by other factors, such as learning and cultural evolution. In some organisms, these causes don't exist at all (bacteria), while in others, they exist but are less powerful (chimps). Further, while trees and bacteria have *apparent* purpose, humans have *actual* purpose (Dennett, 1995). A bacterium scavenges iron from the soil because selection over many generations of those bacteria that scavenge more iron has made it appear as though it wants…. Etc. A human, on the other hand, may pick up a piece of iron because it actually wants to do so. As such, it's usually safer, and clearer, to keep human and non-human explanations as separate as possible to begin with, and later make links.

Cooperation across the tree of life

Contents

- **What is cooperation?**
- **Explanations for Cooperation**
- **The answer lies in the gene**
- **The success of inclusive fitness**
- **The debate over inclusive fitness**
- **Alternative ways to measure cooperation**
- **Cooperation between species**
- **The social amoeba: the perfect test case for cooperation**

What is cooperation?

We'll start with explanations for cooperation that work across the tree of life, for both humans and non-humans -- in other words, *general* explanations for cooperation. To do so, though, we need to define it.

A common criticism you'll hear in science is "that's just semantics", the implication being that the person is quibbling about semantics rather than debating the meat of things. But semantics matter in science. The reason is simple: insofar as it's possible, researchers need to agree on terms so that debate can be fruitful, so that arguments aren't cross-purpose. If I claim that planets can never be within a million miles of a star, and you claim they can, for that debate to be worthwhile, it's important we both mean the same thing by "planet". It might turn out that we actually both agree that a large rock can be close to the sun, and I was simply thinking of that as an asteroid not a planet. Defining our terms, and agreeing on them, ensures that we're all playing the same game, on the same field, with the same rules (West, Griffin, & Gardner, 2007).

So what is cooperation? In evolutionary biology, we can define behavioural traits by the *effects* the trait has. These effects can be on the bearer of the trait, the *actor*, and on other individuals who are also affected by it – the *recipients*. The effects we are interested, are effects on fitness, or *number of offspring*.

Take, for example, a pair of horns as a trait belonging to a big horn sheep, or a ram. We take one ram with the trait, horns, to be the actor. The horns have some effect on the actor's number of offspring. For example, they might cost the ram energy to produce, and have a negative impact on its number of offspring in that way, but they might also win it more mates, and have a positive effect in that way. We are only interested in the *net* effect, positive or negative.

In addition, the horns may impact other rams. For example, they may cause injury to other rams, and reduce their number of offspring, or have a negative effect.

Of course, any trait will have lots of different effects on lots of individuals, but for simplicity, we focus just on the net effect on itself and the net effect on others. To further simplify things, we can think of the net effect on all other individuals as the net effect on a single, average recipient.

This means we can conceptualize all traits by their effect on the actor and a recipient. The four possible types of traits are shown in the figure below (Figure 2).

EFFECT ON RECIPIENT

	+	-
+	MUTUALISM +/+	SELFISHNESS +/-
-	ALTRUISM -/+	SPITE -/-

EFFECT ON ACTOR

Figure 2. Trait Effects Quadrant. Hamilton classified traits by the effect on the actor (the individual exhibiting the trait) and the recipient (the individual, other than the actor, affected by the trait). The left column includes both forms of cooperation.

Selfish traits, those which have a positive effect on the actor, and a negative effect on the recipient, are those which we're most familiar with. Eating up all the food for yourself, for example. These are the traits Darwin noticed first and foremost (*nature, red in tooth and claw*), the traits which are most common in nature, and the traits which are most intuitive to understand. In other words, the boring traits.

We won't talk about spiteful traits here, although they are interesting. Instead, we are interested in the top row of traits – those that have a positive effect on recipients.

That's how we define cooperation, in terms of effect on recipient. More formally, we define cooperation as "a behaviour which provides a benefit to another individual (recipient), and which is selected for because of its beneficial effect on the recipient" (West et al., 2007). This last part is important – in evolutionary biology, we must always define traits in terms of the *purpose for which they were selected*.

Why? The purpose of definitions, beyond scholarly agreement, is to allow us to make clear predictions. In evolutionary biology, we define behaviours, such as cooperation, so that we can make clear predictions about when those behaviours will evolve, why they will evolve, etc.

Now, consider the case of a defecating cow. The cow defecates, and a fly comes and lays its eggs in the poop. The fly certainly benefits from this behavior – the "recipient" is affected positively. But the cow is defecating simply because it is

defecating. The effect on the recipient has had nothing to do with the selective history of that trait. Therefore, if we treated it is as cooperation, and tried to understand why the cow defecates with regards to the fly's benefits, we would be misled.

Thus, we are interested in traits that have a positive effect on others, and were selected because of that effect.

Explanations for Cooperation

Looking at the two types of cooperation in the top row of the figure, altruism and mutualism, obviously one is easier to explain than the other. When a trait has a positive effect on others, *and* a positive effect on the actor, this means that we can explain the behavior, from an evolutionary perspective, by the *direct effect* on the actor. If helping another individual has a net positive effect on yourself, then it can be favoured by selection. For example, when a bee pollinates a flower, it helps the flower, but it also gets food, and we can explain why this evolves because it is directly beneficial to the bee's number of offspring.

This is cooperation via *direct effects* (see Figure 3). It can take a number of forms, and the exact way in which the benefits come directly back to the actor can get pretty interesting. This is *especially* true when the interact organisms have traits like learning, memory, imperfect knowledge – for example, I'll help you, because then, in the future, you might tell Joe, who will remember to help me, etc. For that reason, we'll return to cooperation via direct effects later.
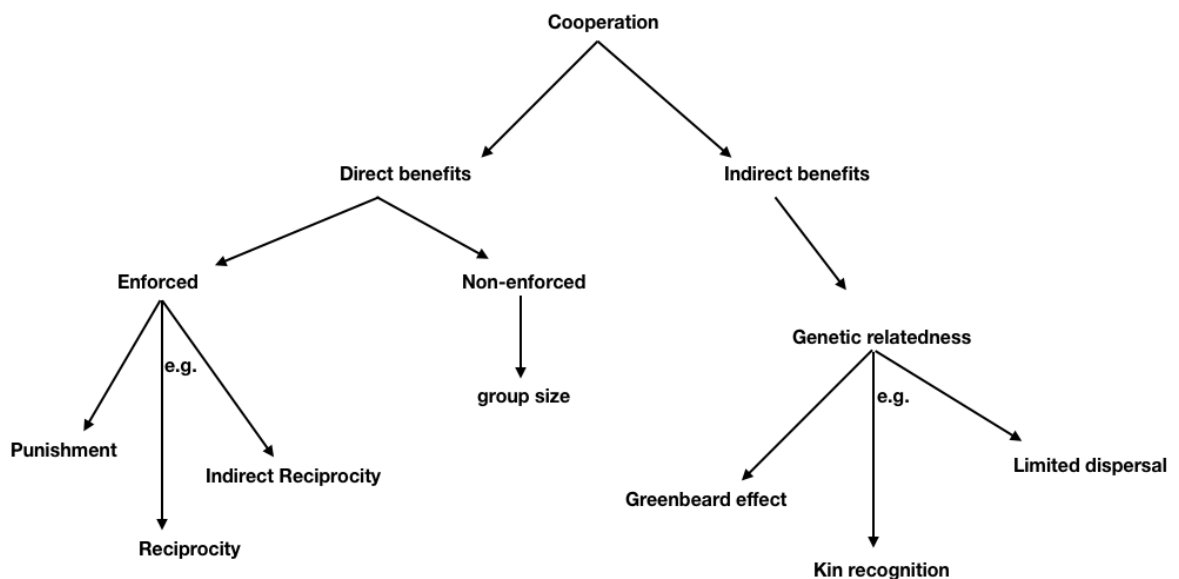

Figure 3. Adapted from West et al. 2007. Cooperation can be explained by indirect benefits, which accrue to gene sharing relatives, or by direct benefits,

which ultimately return to the cooperator. A non-exhaustive list of examples of each is shown here.

For now, we'll focus on the right-hand side of Figure 3. That, of course, is where the plot really thickens, because we return to our question from the introduction: why on earth would organism possess a trait that had a negative effect on itself, but a positive effect on others. How is this possible?

The answer lies in the gene

It was Bill Hamilton who solved this great puzzle. The structure of Darwin's idea wasn't wrong. He thought that organisms should act selfishly to maximize their number of offspring, because organisms that are better at reproducing become better represented in the population. What Darwin didn't know – what he couldn't have known – is that it's not organisms that reproduce, it's genes.

Hamilton showed that because of that, it's *genes* that are better at replicating that become better represented in the population. Genes become 'designed' to maximize their own reproduction. As a result, organisms are made up of genes that are designed to replicate well. Thus, organisms should appear as though they are trying to maximize the spread of their *genes* (Hamilton, 1964).

An organism has two ways to spread its genes into the next generation. It can do so by reproducing itself, and passing its genes to its offspring. Or, it can help other individuals, who share *copies* of its genes, to reproduce. Who shares genes? Relatives. For example, siblings share on average half of their genes, cousins 1/8th, etc. From natural selection's point of view, it is indifferent to whether an individual produces one offspring or two full siblings. Thus, the famous quip, attributed (possibly falsely) to Haldane, "I'd lay down my life for eight cousins or two brothers".

This idea, developed by Hamilton, that a trait can be selected for via its effects on relatives, has been termed Kin Selection (Hamilton, 1964, 1970). But it's not just relatives as you and I are familiar with them – brothers and grandchildren and so on – that share genes. Perhaps Hamilton's greatest insight was that all individuals in a population share genes to a greater and lesser degree.

Thus, selection must count the effects of a trait not just on its bearer, the actor, but on all individuals in the population. Selection should weight the effect on the actor most strongly, because the actor shares all its genes with itself, and a full sibling about half as much, on and on, to a distant relative who might be valued at about 0.1. The degree to which individuals share genes, or are *related*, can be measured precisely, and it's this value of relatedness that selection should use to weight the effects of a trait (Alan Grafen, 1985; Hamilton, 1964, 1970).

Hamilton formalized this notion with a simple rule, which tells us whether a trait will be favoured by selection. Count the effect on the actor, or cost C, and add to it the effect on all the recipients in the population, or benefit, B. But an effect on a sister should be weighted by ½, and cousin 1/8, and any random individual by their exact relatedness, or probability of sharing genes, R. If the sum of the costs and benefits is greater than 0, the trait will evolve:

$-C + RB > 0$

This is known as Hamilton's rule, and it can tell us whether any trait will evolve. For example, consider a trait which costs the actor 2 offspring, but gives it's sibling, related by ½, an extra 10 offspring:

$-2 + 0.5*10 = -2 + 5 = 3 > 0$

Thus, this trait should be favoured by selection. This idea can be recast in economic terms, too: imagine you were trading US dollars (USD) for Chinese yuan (RMB). You'd be willing to trade, or give up, some number of dollars, C, for some number of yuan, B, given those yuan, weighted by devaluation, outweighed your dollars. In that case, the appropriate devaluation is the international exchange rate. In the case of biology, an organism will give up its own offspring in exchange for other individuals' offspring, where the exchange rate between them is relatedness, R (Frank, 1998).

Hamilton captured this idea more broadly in a measure he called "inclusive fitness". In contrast to simple fitness, which only counts offspring of your own, inclusive fitness measures both the offspring you have yourself, and those you are responsible for that others have, where those foreign offspring are devalued by your relatedness to them(Frank, 1998; Alan Grafen, 1985; Hamilton, 1964, 1970).

It's hard to overstate how huge these two idea were (Queller, 2016b, 2016a). The first, kin selection, extended Darwin's theory to explain how selection acting on genes impacts individuals. The second, inclusive fitness, extended Darwin's theory to identify the new ultimate purpose of organisms. Darwin knew nothing of genes, and thought organisms should be designed to maximize their number of offspring. Hamilton showed that, because of genes, organisms should appear designed to maximize their inclusive fitness, or the number of offspring they have *plus* positive effects on relatives' number of offspring (Alan Grafen, 2006, 2015, Hamilton, 1964, 1970; Queller, 2016a).

Of course, Hamilton's theory also explains cooperation. In fact, Hamilton's simple rule above, explains all four types of traits in the trait quadrant:

EFFECT ON RECIPIENT

		+	-
EFFECT ON ACTOR	**+**	MUTUALISM +/+ $-c + rb > 0$ $+ \, + r+ > 0$	SELFISHNESS +/- $-c + rb > 0$ $+ \, + r- > 0$
	-	ALTRUISM -/+ $-c + rb > 0$ $- \, + r+ > 0$	SPITE -/- $-c + rb > 0$ $- \, + r- > 0$

Figure X. Trait quadrant with Hamilton's rule. Hamilton's rule says when a trait can evolve. If rb is greater than c, the trait can evolve.

Altruism, cooperation which is costly to the actor, can be explained if the benefits to recipients, weighted by their relatedness to the actor, outweighs the costs. This is cooperation via *indirect benefits*, because the benefits accrue not to the actor, but to its gene-sharing relatives.

At last, we can finally can explain the sacrificial worker bee. A bee colony is a large family. The queen is the mom, and the sterile worker bee shares genes with any offspring produced by her – they will be the worker's siblings. When the worker stings a predator to defend the nest, it may be dying, giving up its own chance to reproduce. But if doing so saves the colony, and allows the queen to produce hundreds, or thousands more siblings, the trade is a no-brainer. The worker gives up a few offspring of its own to save thousands of individuals, all sharing roughly a half of the worker's genes (in bees the genetics are actually a little more complicated than that, but that's a story for later).

Thus, cooperation via indirect benefits can be explained as long as the benefits accrue to relatives (and, *fundamentally*, that those benefits, weighted by relatedness, outweigh the costs it's important to remember the Bs and Cs as well as the R). But importantly, the benefits have to accrue *preferentially* to relatives. Imagine a trait that was costly to the actor, but gave everyone else in the population two extra offspring. This would give extra offspring to relatives, but it would give *the same number* of extra offspring to non-relatives. This effect would dilute the benefits to relatives, and exactly cancel out.

If that's confusing, try to envision the fate of a mutant gene that helped everyone in the population equal. Let's call this mutant gene A*, to distinguish it from the regular gene in the population, A. In the next generation, there would be extra copies of A* as a result of the relatives who the actor helped reproduce, who also had a copy of A*. But there would also be extra copies of A, from all the non-relatives that the actor helped reproduce. Thus, A* would not have increased in frequency at all.

To solve this problem, relatedness is measured *relative to random member of the population* (Alan Grafen, 1985). The benefits have to go to relatives *more* than they do to everyone else. How might this happen?

Here are some examples:
1. Limited dispersal of offspring. If offspring stay near where they are born, then individuals will tend to find themselves near relatives. Imagine a tree that drops its seeds right down to the ground. A tree that expressed a cooperative behavior, like "release protective antifungals into the soil" would inevitably be helping relatives more than non-relatives, because all its neighbors will be siblings.
2. Kin recognition. An organism might use a cue, such a smell, to preferentially target their cooperative behavior to relatives
3. Greenbeard genes. A special kind of kin discrimination in which a gene codes for a recognizable trait (e.g. a green beard), as well as a cooperative behavior targeted towards individuals with green beards.

Any of these mechanisms, plus many more, can lead the benefits of cooperation to fall preferentially on relatives. This generates a positive R term in Hamilton's rule, and, given RB is greater than C, the trait can evolve.

The success of inclusive fitness

For fifty years, since Hamilton's first set of papers on the idea, inclusive fitness and kin selection have spawned a vast body of theoretical and empirical work, creating the field known as "social evolution" (Davies, Krebs, & West, 2012; Foster, 2009; Krebs & Davies, 1978, 1987, 2009; Queller, 2016a; Westneat & Fox, 2010).

Amongst the leaders in this field are Joan Strassmann and David Queller at Washington University Saint Louis, who have been pioneers both of extending the theory of inclusive fitness (e.g. Liao, Rong, & Queller, 2015; Queller, 1985, 1992, 1994, 2014, 2017; Queller, Rong, & Liao, 2015) and devising ingenious experiments and field studies to test its predictions (e.g. Buttery, Smith, Queller, & Strassmann, 2013; Gilbert, Foster, Mehdiabadi, Strassmann, & Queller, 2007; Inglis, Ryu, Asikhia, Strassmann, & Queller, 2017; Kuzdzal-Fick, Fox, Strassmann, & Queller, 2011; Ostrowski et al., 2015).

Strassmann and Queller summarize the empirical successes of inclusive fitness in a recent review (Queller & Strassmann, n.d.), which range from the origin and distribution eusocial insect colonies, cooperation in microbes, sex ratios in insects, cooperative breeding in birds, to a wide range of other biological phenomena.

One of the most fascinating subfields of research to emerge from social evolution is the study of major transitions in individuality (J. M. Smith & Szathmary, 1995; West, Fisher, Gardner, & Kiers, 2015).

Life started with simple, replicating molecules. And yet we are complex organisms, made up of trillions of cells, which themselves are made up of nuclei and mitochondria, which in turn consist of many thousands of genes, or replicators. Thus, the history of life has involved the coming together of independent organisms to form new individuals. For example, at one point all living things were single-celled organisms, living, eating, competing. But at some point some of these single celled organisms became mere parts of a whole new organism, no longer individuals in their own right, but merely parts of a larger being. These events – when individuals come together to form a new type of individual – are known as major transitions (J. M. Smith & Szathmary, 1995; West et al., 2015).

The major transitions that have happened on earth include events when a single species formed a new organism, such as when single celled individuals became multicellular, and when two species came together to form a new individual, as when two types of bacteria formed the eukaryotic cells that make up you and me (one becoming the nucleus and the other becoming the mitochondria). These two types of transitions are named fraternal (between brothers), and egalitarian (Queller, 1997).

Major transitions are the most extreme form of cooperation. They require individuals that were once competing to subjugate their own interests to such a degree that they become mere parts, or cogs, in a new individual. Consider the cells in your hands in your heart, which are not individual organisms, but simply tools to help your germ line, or the worker ants, bees, and wasps in eusocial insects colonies, which are sterile, merely helping the queen reproduce.

Are the two types of transitions – fraternal and egalitarian -- the same? Not exactly. Fraternal transitions involve individuals of the same type, within a species. As such, they can be explained by individuals sharing genes. The cells in your hand and heart are clones of your sperm or egg cells. Multicellular organisms tend to start

as a single cell, meaning all subsequent cells are copies. Thus, in fraternal transitions, some individuals can give up reproduction altogether, because they are helping their relatives who share genes (Queller, 1997).

Egalitarian transitions evolve two different species, such as a fungus and an algae in lichen. If one partner gave up reproduction, it would disappear altogether! These bonds are maintained by other mechanisms, which we'll talk more about later.

A number of theoreticians have investigated the different types of transitions, and what they have in common and what makes them different (Gardner & Grafen, 2009; Queller, 1997; J. M. Smith & Szathmary, 1995; van Gestel & Tarnita, 2017; West et al., 2015). The Martin Nowalk group at Harvard, along with Corina Tarnita of Princeton, have developed mathematical models to explore the difference between "coming together" and "saying together" (Tarnita, Taubes, & Nowak, 2013).

"Staying together" occurs when direct descendants physically associate to form the group. For example, in the origin of multicellularity, this would have involved a single cell making a clone of itself, and that clone sticking with the original. That process could repeat until you had a clump of cloned cells.

"Coming together" involves to disparate individuals forming an aggregate. For example, the origin of the eukaryotic cell likely involved one bacterial cell engulfing another, different type of bacteria.

In a series of mathematical papers, Nowak et al. show that these two different processes involve very different evolutionary challenges, and can solve different kinds of evolutionary problems (Ghang & Nowak, 2014; J. W. Olejarz & Nowak, 2014; Tarnita et al., 2013).

For example, staying together allows the different units to 'divide up labour' and specialize in tasks, but all the different specialties have to come from a single collection of genes. In our bodies, your muscle cells can do muscle work and your neurons can do brain work, but both designs have to be programmed by the same genome (Tarnita et al., 2013).

Coming together can allow individuals with vastly different skills to combine to achieve something very new. Consider, for example, the partnership between aphids and buchnera bacteria. Aphids allow the bacteria to live in their gut, eating the food the aphids eat, and protecting them. The bacteria synthesize essential amino acids that the aphids can't get otherwise. Neither partner can live without the other. Bacteria could never produce a mobile home as good as an aphid, and aphids can't change their metabolism enough to synthesize new amino acids. However, the two partners are limited in just how much they can specialize, because both partners need to remain functional enough to reproduce. These ideas are old ones (Queller, 1997), but are being given, here, a formal mathematical treatment (Tarnita et al., 2013).

The two processes invoke different challenges, too. Staying together means that relatedness has to be high enough that some individuals are willing to give up reproduction (Queller, 1997). This can be resolved by clonality (multicellularity) or strict lifetime monogamy (eusociality) (Hughes, Oldroyd, Beekman, & Ratnieks, 2008; West et al., 2015). Coming together becomes an evolutionary game, in which both partners need their interests to be sufficiently tied together. In the case of aphids and buchnera, their reproduction events are inextricably linked.

An exciting future direction for this work will be to see whether we can develop a fully general model that predicts when a major transition will happen, egalitarian or fraternal, or whether we need separate models for the different types of transition.

More generally, inclusive fitness theory, and in particular the major transitions research programme, have revealed that cooperative behavior, once viewed as a troublesome aberration in the theory of natural selection, is both readily explained with evolutionary biology, and a crucial component of all life. Queller and Strassmann (Strassmann & Queller, n.d.) have termed cooperation the "most underestimated process in the evolution of life." One of the satisfying features of the theory is that it unites cooperation across the tree of life under a single explanatory framework (A. Aktipis, 2016; West et al., 2015; West & Gardner, 2013).

However, despite the enormous success of Hamilton's theory, the idea of inclusive fitness has been criticized since at least 1978 (Cavalli-Sforza & Feldman, 1978), and these criticisms have become increasingly heated in the last decade.

The debate over inclusive fitness

In 2011, nearly 140 authors wrote a reply to a paper that had been published in Nature, arguably the most highly rated journal in science, the year before. This was a rare event in the world of academia.

A year earlier, in 2010, nearly 50 years after Hamilton's original papers on inclusive fitness, three authors had published a paper criticizing the idea (Nowak, Tarnita, & Wilson, 2010). The authors, all based at Harvard at the time, were E.O Wilson, a giant of biology and entomology, Martin Nowak, a leading mathematical biologist, and his student, Corina Tarnita (now an associate professor at Princeton).

While debates over inclusive fitness had occurred for many decades, the original Nature paper and its subsequent replies and replies sparked a particularly charged exchange which has lasted now for ten years, at times becoming unusually political for scientific discourse.

The 2010 paper was a mixture of verbal and mathematical arguments. The verbal arguments claimed that inclusive fitness had hindered progress in empirical work, handicapped theoretical work, had little to no empirical support, was impossible to calculate, and offered no added insight to what they call "standard evolutionary dynamics". The mathematical part claimed to show, formally, that inclusive fitness gives the same "answer" as other, simpler, models, and that relatedness, a key factor in Hamilton's theory, doesn't matter (Nowak et al., 2010).

The paper was soon followed by a reply, authored by nearly 140 researchers from around the world, refuting the paper (Abbot et al., 2011). The response claimed that inclusive fitness was founded on standard evolutionary theory, and was as general as the standard theory of natural selection, and therefore not surprising that it made the same predictions. Further, they included an expansive list of the many empirical applications of inclusive fitness, claiming that Nowak et al. had ignored a huge body of research ((Abbot et al., 2011)).

Many subsequent papers have followed, from both sides of the table. Two particularly interesting episodes in the debate are a 2013 paper by Wilson, Nowak, and a third colleague, Benjamin Allen, and a series of papers by David Queller and several students, from 2015 to 2016.

The 2013 paper by Allen et al. focused on certain assumptions made in mathematical models of inclusive fitness, distinguishing between simple forms of inclusive fitness models, which make a number of assumptions and are therefore easy to interpret and apply, and more general models, which make few assumptions, but are harder to interpret. They claimed to show, mathematically, that the simple models are never true, and the complex models have no meaning (Benjamin Allen,

Nowak, & Wilson, 2013). Others have argued that this was a misinterpretation (Birch, 2014; Birch & Okasha, 2015).

Meanwhile, the paper by Queller and colleagues claimed to show that Nowak et al. (2010) misinterpreted their own mathematics. Liao et al. reanalyzed the original model, extending it to consider a wider a range of possibilities, and claimed to find that, after all, the analysis shows that relatedness *does* matter (Liao et al., 2015). This sparked a reply by the Harvard group, and a subsequent reply by Queller and his colleagues (Queller et al., 2015).

This finaly reply was titled "Some agreement on kin selection and eusociality?". Perhaps that is where the debate is headed. Interestingly, the arguments have become increasingly mathematically technical, making it difficult for many empirical biologists to make sense of the claims and results. Perhaps this means biologists can carry on with their day jobs while the theoreticians debate. Or perhaps, eventually, the gap between the mathematical arguments on one hand, and biological practitioners on the other, will be bridged, and some resolution will be reached.

A notable exception to the highly mathematical arguments is Queller's 2016 paper, "Kin selection and its discontents" (Queller, 2016a). This is certainly one of the clearest, and most thorough defenses of Hamilton's ideas. Queller outlines 15 claims that have been made by critics of inclusive fitness, and responds to each in turn. The overall message is that kin selection and inclusive fitness theory have been enormously successful, and that the vast majority of criticisms have come from misunderstanding.

Alternative ways to measure cooperation

Researchers from both sides of the debate have always acknowledged that inclusive fitness theory isn't the only way to understand cooperation. Martin Nowak of Harvard has been a champion and pioneer of these alternative measures, the most notable of which is the use of graphs to understand evolution.

Hamilton's insight was that whether a gene spreads depends not only on the gene's carrier, but on the effects that gene has on other individuals, as well. There are two obvious reasons: 1. There may be copies of that gene in other individuals, so whether the carrier helps or harms those neighbours matters for how common the gene is in the next generation. 2. The frequency of a gene is what matters (being in 8 individual's out of 10 is more valuable than being in 10 out of a million), and frequency is relative. Therefore, regardless of whether other individuals share your gene copies, your impact on their offspring number will matter for your gene's representation in the next generation.

Hamilton incorporated this fact into his models through relatedness, a measure of the degree to which neighbours share genes, and costs and benefits, measures of the fitness impact neighbours have on each other.

But another way to measure these factors is by modelling evolution on a "graph" (Ohtsuki, Hauert, Lieberman, & Nowak, 2006). What this entails is imagining that there are a bunch of "sites", or locations, where animals can be – a little like the black and white tiles on a chessboard, or the hexagons on a Catan board. These sites can be isolated or connected, to any number of other sites. The connections determine how animals move between the sites, and what kinds of interactions can happen between them. In chess, for example, you have four immediate neighbors, while in Catan you have six. In chess, a bishop stays on the same colour tile the

entire game, and a king can only move one tile at a time. Similarly, different graphs have different arrangements of sites, determining how many neighbours individuals have, and different connections between them, determining how individuals move about. In the example below,
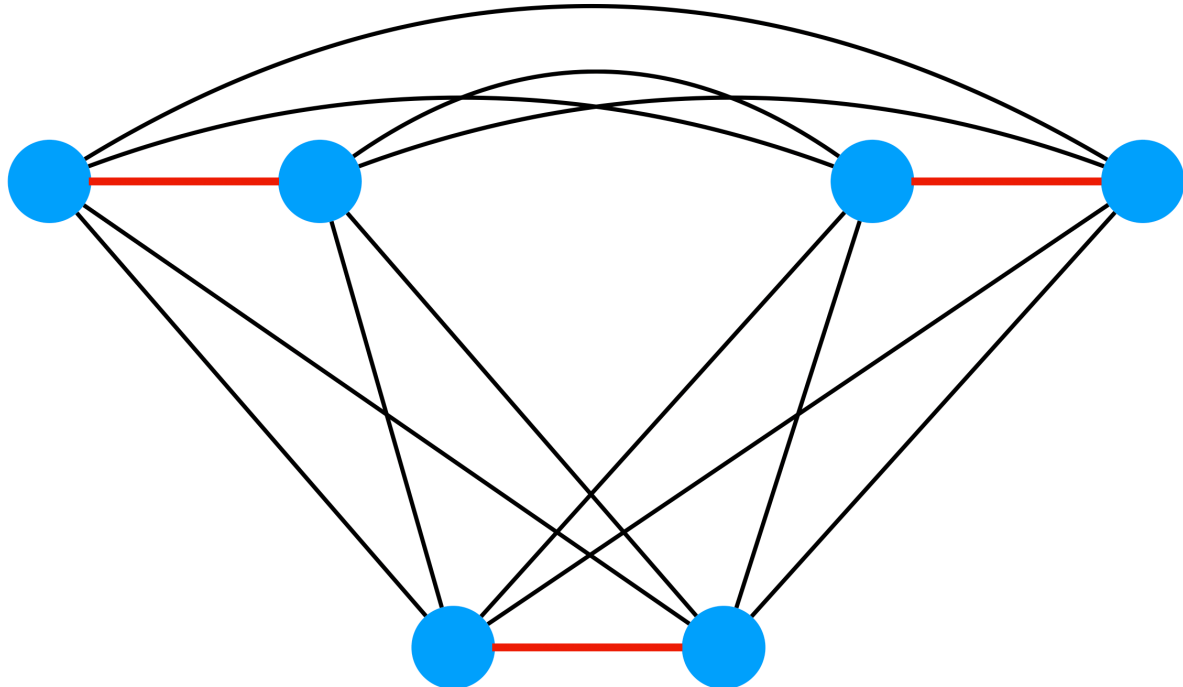
once you have a graph, you can place a population of organisms on it. If you are interested in cooperation, you might allow organisms to be cooperative or selfish, and allow them to interact, affect each other's fitness, reproduce, etc., and ask whether cooperation spreads, or selfishness spreads, or some stalemate is reached.

This is what Nowak and colleagues have done, for a number of different graphs and interactions. In one example, they used this approach to model public good-producing bacteria.

Public goods are products that individuals work to produce or acquire, but that are shared collectively by the group. For example, some bacteria release (expensive) chemicals which help free up iron in the soil, a valuable resource. The cost of producing the chemicals is shouldered by the individuals who release the iron-scavenging chemicals, but the benefits – the free iron in the soil – are shared by everyone in the neighborhood.

The question for Nowak and his colleagues was: why would individuals incur the cost of making a public good (Benjamin Allen, Gore, & Nowak, 2013)? A good strategy might be to cheat, and not produce the expensive chemical, but partake in enjoying the public good. In other words, why cooperate?

Allen et al. modelled the bacteria either on hexagonal lattices (like in Catan) or on circular, connected sites. They found that several factors determined whether cooperation evolve: the costs and benefits of the public good, the degree to which the good is private (the cooperator keeps some of its own good produced), the degree of connectedness of the graph, the decay rate of the public good, and the diffusion rate of the good. They showed, mathematically, that lower diffusion rates, lower connectivity of the graph, and slower decay of the public good favoured cooperation (Benjamin Allen, Gore, et al., 2013).

In some ways, this result might seem intuitive – the less you have to share your public good, and the longer the benefits of making it can return to *you*, the more likely you are to expend the effort to produce it (e.g. to cooperate). On other hand, being able to demonstrate these results mathematically, using simple graphs, is an exciting theoretical achievement, and allows us to make future predictions which can be tested experimentally.

After studying specific biological problems like public goods on graphs, Nowak and colleagues moved on to more general issues, trying extract very general, universal principles of evolution from the graphs they stated.

This work is highlighted by two key results from their team in 2013 and 2014. In 2013, Allen et al. showed, using the mathematical graphs approach, that just two features of a biological system can be used to understand and predict its evolutionary behavior (Benjamin Allen, Nowak, & Dieckmann, 2013).

The first is what they call a "structure coefficient" essentially captures the way in which the sites on the graph are connected, and how individuals change sites, within and between generations. The second is what they call "effective population size", which is a measure of how quickly a beneficial mutation will spread and become fixed in a population.

Allen et al. find that these two numbers are, in many biologically relevant scenarios, sufficient to make evolutionary predictions about how and when traits will spread, independent of what kinds of traits we are studying (Benjamin Allen, Nowak, & Dieckmann, 2013)! This is a remarkably general result.

Next, in 2014, Adlam and Nowak extended the reach with which graphs could be used to explain the natural world. One limitation of using graphs as models, is by forcing organisms to live on sites and only move along certain paths and interact along rigid connections, they vastly oversimplify nature. Of course, any model simplifies nature – that's the point of a model! But one hopes that they don't lose *too much* relevant information about how the world works in the process. In the case of graphs, the concern is that by making the structure of a population rigid, important information about evolution might be lost.

However, Adlam and Nowak showed, mathematically, that the predictions from rigid graphs *aren't so different* from predictions using a fully well-mixed population (Adlam & Nowak, 2014)! The key result was that, as graphs get bigger, they effectively become the same as well mixed populations. This provides us with some confidence that we can use these oversimplified population structures, and therefore make our calculations simpler, without getting our picture of the world too wrong.

Therefore, graphs have joined the arsenal of tools for understanding evolution, and, in particular, cooperation (Benjamin Allen & Nowak, 2014). An astute reader may have noticed a connection between the idea of connectivity in graphs, and relatedness in inclusive fitness theory. Indeed, researchers have shown that relatedness can be calculated on a graph, and that inclusive fitness models and graph models give identical results (A. Grafen, 2007). This is always incredibly satisfying in science – when very different approaches give the same answer, it suggests we are on the right track. Exactly how well these approaches line up, and the precise nature of their relationship, is an active area of research (e.g. Taylor, 2017).

Cooperation between species

So far, we've focused on cooperation between members of the same species. A bacterium helping another bacterium, a honey-bee helping another honey-bee. This kind of cooperation can often be explained through indirect benefits – the sharing of genes – which we have focused on.

But what about the cooperation between a honey-bee and a flower? This type of cooperation can't be explained through indirect benefits – after all, bees and flowers aren't related. They don't share genes[1].

We will talk more in depth about cooperation via direct benefits in the section on human cooperation, because such a vast body of research is dedicated to that topic. But it's worth, here, saying something about how we understand cooperation between non-human species.

As mentioned earlier, Ronald Fisher, both the father of modern statistics and modern evolutionary biology, provided formal support for Darwin's theory of design. His mathematics were complicated and poorly explained, but the simplest way to understand his result is that he showed that selection causes fitness to increase (even while other forces may have other effects). One interpretation of this result is that it proves that, in the absence of social interactions, over time organisms will consist of traits that increase their personal fitness, or offspring number

Fisher's main equation has come to be known as The Fundamental Theorem of Natural Selection. David Queller has eloquently argued that, despite this equation's importance, it actually deserves a small demotion (Queller, 2017). The true fundamental theorem, he argues, is Price's equation, from which Fisher's equation is derived. Price's equation, for which the mathematics are equally formidable and the derivation equally opaque, states simply that the change in a trait over time is equal to the covariance (a statistical relationship akin to correlation) between that trait and fitness. In other words, fitness determines how traits change over time.

But Queller has taken this idea one step further. Traits, he points, are often not just properties of a single organism, but multiple (Queller, 2014). For example, the portion of a leaf eaten by an aphid is a property of both the plant and the aphid. Queller used this idea to extend the Price equation, the fundamental equation of selection, to allow the trait under selection to be a property of more than one organism.

This result is critical for conceptualizing how we expect evolution to proceed when two species are involved. If a trait has a negative effect on fitness for two species, we expect it to disappear. If it has a positive effect on fitness for both species, we expect to spread, as might be the case with the amount of visits a bee makes to a flower. Things get especially interesting when the fitness effects are opposite, as in the case of the leaf and the aphid – the more of the leaf eaten, the better for the aphid and the worse for the plant. Who wins these conflicts?

Even in the case of the flower and the bee, there may be some conflict. The flower wants the bee to spread its pollen, but for a minimal amount of nectar doled

---

[1] Actually, they *do* share genes. All life is descended from a common ancestor. After all, most of our genome is shared with a chimpanzee. But remember, what matters to selection is the likelihood of sharing genes *relative* to the rest of the population. The genes a flower and a bee share are also shared by that flower and all other bees. Helping the flower reproduce doesn't increase that gene's representation relative to not helping.

out. How does this potential conflict get resolved? Queller's equation for joint-species traits provides us with a clear, formal architecture for studying these traits (Queller, 2014).

When do two species cooperating, such as the flower and the bee, become so integrated that we no longer think of them as separate species? Lichen, for example, is a composite of a fungus and a bacteria or algae. But for a long time people assumed it was a single species. Or take the very cells in our bodies – your mitochondria and nucleus were once fully separate, independent species. Can we still think of them that way, or as a single species, or something in between?

In an eloquent perspectives piece, Joan Strassmann and David Queller explore this question of organismality (Queller & Strassmann, 2016). The concept of an organism, they point out, is fundamental to biology, and yet there are border cases where it's not clear what an organism is.

Clearly the honey bee and the flower are two separate organisms. And few would argue that the eukaryotic cell is two species – the mitochondria and the nucleus have been joined for nearly 2 billion years, and have long since lost any ability to function independently of each other. Is there a clear line between the two? What about aphids and buchnera bactieria, which live in their gut? The aphids can't acquire essential amino acids, and would die without the aphids, and the bacteria can't live outside the gut of the aphids. Are they a single organism?

Queller and Strassmann argue, following the principles highlighted in Queller's model (Queller, 2014), that the answer lies in conflict. There are still conflicts of interest between the flower and the bee, as mentioned above. The flower wants the bee to take less, and the bee wants to take more. These conflicts prevent selection from moulding the two species into a single, fully cohesive entity. In the case of the mitochondria and the nucleus, they reproduce in unison. If either one gets more copies into the next generation, the other gets more copies into the next generation, too. This means there are effectively *no conflicts of interest* between the two. This allowed selection to shape them into one unit, to the point where they were completely *mutually dependent*.

Queller and Strassmann say this is what defines an *organism*: an entity with effectively no internal conflicts (Queller & Strassmann, 2016). The word effective is important, because even an organism like you, has some conflict. Consider, for example, cancer, which can be thought of as one cell trying to replicate at the expense of the rest of the organism (Athena Aktipis et al., 2015). But these conflicts are limited as they can't spread into the next generation. Thus, when two cooperative organisms shed all conflict, they can become a new, single organism.

The social amoeba: the perfect test case for cooperation

We have, so far, talked mainly about *theory*. Theory is an essential part of any field of science. It's where we articulate our understanding of the world in the form of statements, models, hypotheses -- all of which make *predictions* which embody our conception of reality as it stands currently. In most mature sciences, this theory often takes the form of mathematics, in part because math goes some way towards resolving ambiguities and gaining precision. This has become increasingly true in biology, exemplified by David Queller and Martin Nowak's work described above.

But theory is only one side of the scientific coin. The other side is empirics, where the predictions from theory are tested. Empirical work can be done in the field, in the lab, or through comparative work. Many subfields of biology have, historically, converged on a 'model-organism' – an organism that possesses certain traits that lend itself to being studied either in the lab or the field. In genetics, the fruit fly is a good example, used widely because it is easy to keep alive in the lab, and breeds quickly. If researchers are lucky, they find the perfect organism for studying the specific problem they are interested in.

In the field of social evolution, one would be hard pressed to give this mantle to anyone other than the "social amoeba". The "social amoeba" usually refers to *Dictyostelium discoideum*, or other members of *Dictyostelium,* a family of slime moulds.

In the case of *Dicty*, it as though the skies open up and delivered social biologists a model organism on a platter. And Joan Strassmann and David Queller, of Washington University, Saint Louis, have spent the last decade exploiting this opportunity with ingenious experiments.

A pseudo-organism

One of the claims of kin selection theory is that it can explain both the origin of multicellularity and certain common features of multicellular organisms. Multicellularity is an example of *extreme cooperation*. The cells in your body, the distant descendants of one free-living, single-celled organisms, are now so integrated they appear mere parts of a larger individual. They divide up labour, specializing in different tasks, and exhibit total sacrifice; the vast majority of cells in your body don't reproduce from one generation to the next. They sacrifice themselves for the few germ-line cells that do.

The prediction of kin selection theory is that this cooperation is explained by the extremely high relatedness between the cells in a multicellular body. In you or I, the cells in our body our clonal. Therefore, they should be indifferent to whether they reproduce or another cell in the body reproduces, as they share the exact same genes. Conflict of interest is absent. Further, kin selection theory predicts that this is why most multicellular organisms start life as a single cell – e.g. the zygote – because doing so results in a fully clonal organism.

This is a really neat prediction – but one that is nearly impossible to test. The reason is that nearly all multicellular organisms *are* started with a single cell. It would be nice to compare, for example, between multicellular organisms that were started with one cell (clonal relatedness), two cells, (slightly less than clonal), etc., and see whether the less clonal organisms had more evidence of cheating, conflict, selfishness, and so on.

Or, even better, it would useful to take a multicellular organism, like a mouse, and whether it starts as a single-celled zygote (maximal relatedness), or is founded by the coming together or multiple unrelated cells, and see what impact this has on cooperation. But mice, and most other multicellular organisms, can't be started with two or three cells. They can only be started with a single zygote.

In comes Dicty. The social amoeba is a rare organism that can be both single celled and multi-celled. It exists as a single celled amoeba, crawling around the soil eating bacteria. But, when the amoebas begin to run out of food, or starve, they can aggregate to form a multicellular "slug". This slug then crawls along the soil to a good anchoring location, at which point a remarkable transformation begins to take place.

Some fraction of the cells in the slug, roughly 20%, form a vertical stalk jutting up from the ground. To do so, they have to sacrifice themselves, and die, forming rigid stalk material. The remaining 80% of cells climb up the stalk where they become reproductive spores. From there, they can be dispersed, for example by wind, to a new locations, where they start the life cycle over as single celled organisms.

Figure X. *Dictyostelium discoideum*. Photograph by Usman Bashir, licensed under creative commons.
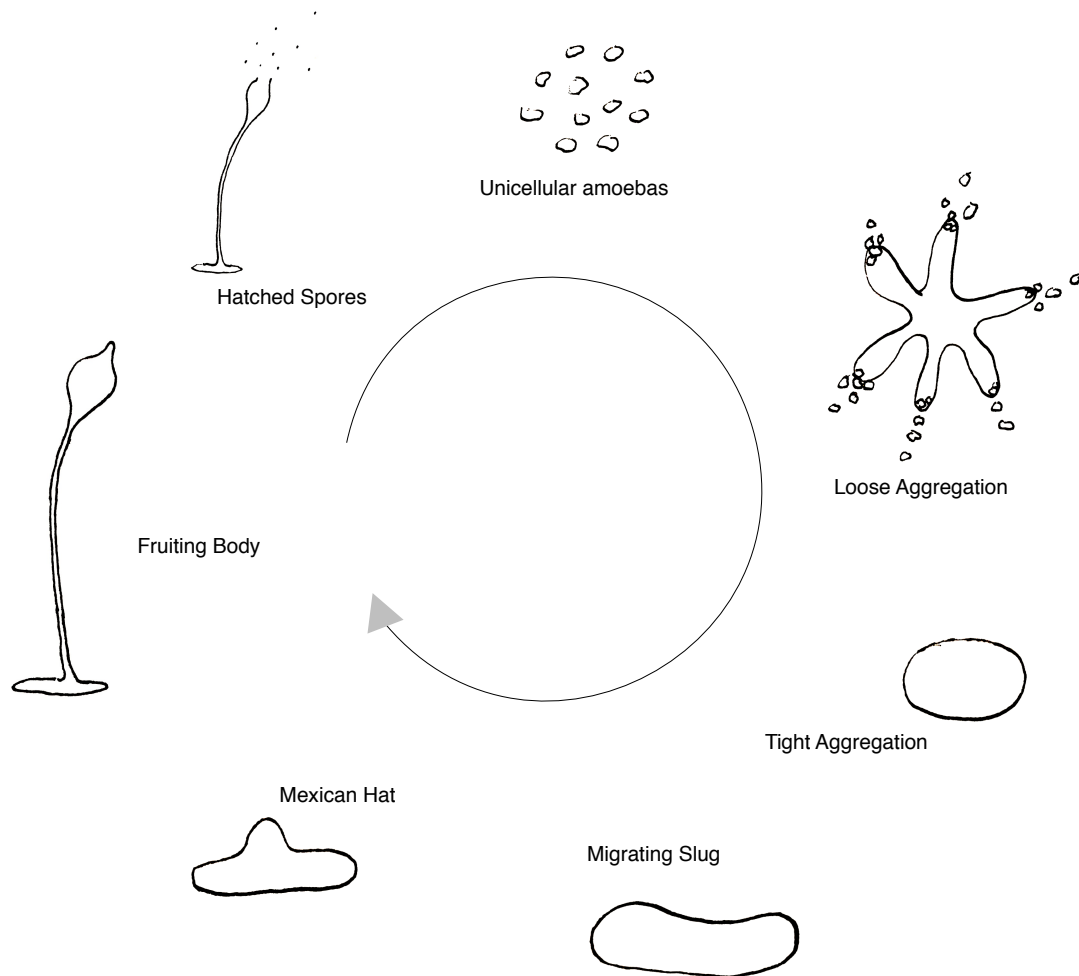
Figure X. Life cycle of *Dictyostelium discoideum*. Single celled amoebas, when starved, aggregate to forma migrating slug. The slug, after migration, forms a fruiting body, with ~20% of cells sacrificing to form the dead stalk, and the other forming the reproductive spore head. These spores are carried to a new location where the life cycle restarts.

It's hard to overplay how wild this is in comparison the vast majority of multicellular creatures. First, their single-celled stage is free-living. This would be akin to our zygote spending a couple of years wandering around eating and dividing. Second, their multicelled stage is *facultative*, meaning they may or may not enter it (unlike ours which is *obligate*). Third, despite the cells not being clonally related, twenty percent of them sacrifice for the other eighty – why? How could natural selection favour this? But perhaps most exciting of all, these features together mean

that scientists have a way to test all the kin selection theories about multicellularity. Which is exactly what Strassmann and Queller have done.

First, Strassmann and Queller wanted to test the prediction that high relatedness explains cooperation in multicellular organisms. To do this, they invented something they called the "pseudo-organism" approach (Queller & Strassmann, 2013). Because Dicty's single celled stage was free living, and because the multicellular organism was founded by more than one individual, they could experimentally manipulate almost all the features of the multicellular cell organism: how many cells founded a slug, how closely related those cells were, how they got chosen to *be* in the slug, etc. This meant that all of the predictions that were previously untestable, were now accessible to experiments.

Does high relatedness explain cooperation?

The first of these experiments was to test whether high relatedness explains cooperation. To achieve this, Strassmann, Queller, and their team created a number of distinct lineages of multicellular amoebas, with varying degrees of internal relatedness. They grew amoebas on petri dishes, feeding them with bacteria. As amoebae grow and divide, natural mutations occur, meaning that the petri dish would contain a significant amount of genetic diversity (i.e. non-relatives). When the amoebas ran out of food, they would form the multicellular fruiting body. In order to start the next "generation", the experimenter could either grab the head of a single fruiting body, or a number of fruiting bodies, or all the fruiting bodies on the petri dish, and then plate these spores out on new petri dishes.

Whether they grabbed one or a thousand would determine how closely related the subsequent amoebas would be. In the extreme case of a single spore, this is akin to a single zygote founding a multicellular organism. In the case of many thousands, this could result in extreme low relatedness.

Of course, selection can't shape adaptations in a single generation. Thus, the researchers repeated this process for many rounds, or "pseudo-generations" of fruiting bodies. In some lines, they always took only a very tiny number of spores to start the new generation (high relatedness). In others, they always took a big swath of fruiting bodies, meaning there was always high genetic diversity (low relatedness). The prediction, of course, being that in high relatedness lines cooperation would be maintained, and in low relatedness lines cooperation would be reduced or even lost altogether.

The predictions were born out by the results (Kuzdzal-Fick et al., 2011). In the high relatedness lines, after thirty-one generations, the cooperative social amoeba remained little changed. However, in the low relatedness lines, a number of cheaters emerged. These cheaters acted in two ways. Some cheaters simply increased their representation in the reproductive head of the fruiting body, contributing less than their fair share to the sacrificial stalk. But others gave nothing at all to the stalk, becoming purely selfish. These cheaters are so extreme that they lost the ability to form sacrificial stalk cells at all. This means that when grown with cooperative individuals, they get all of their cells into the reproductive spore. But when grown on their own, they can't form fruiting bodies at all, because none of them are "willing" to act as stalk.

This result lends strong support to the idea that high relatedness maintains cooperation in multicellular organisms. It suggests that in the absence of high relatedness, multicellularity would be difficult to evolve. But it raises a further

question. Dicty had already long since evolved cooperation when this experiment began. Given that cooperation was present, as relatedness began to decrease, and the selection pressure for cheating increased, do the cooperators have any ability to evolve resistance to that cheating?

Queller, Strassmann, and their team went on to answer this question, using the lineages evolved in the previous experiment (Levin, Brock, Queller, & Strassmann, 2015). They focused on the low relatedness lines, in which cheating evolved. Importantly, in the thirty-one generations over which these cheaters evolved, there was also plenty of time for evolution to act on the non-cheaters (cooperators) as well. If those cooperators evolved any ability to "resist" cheating – in other words, to limit the cheaters success at getting into the reproductive spore head – then they should behave differently to the original, unevolved ancestor of all the lineages.

Fortunately, the spores of amoeba can be frozen and unthawed. This means that the ancestor of all the lineages in the Kuzdzal-Fick experiment was waiting on ice. The prediction of Strassmann and colleagues was this: if resistance has evolved amongst the cooperators, they should *perform better against cheaters than their ancestor*. In this case, performing better means doing a better job at keeping the cheaters out of the spore head.

Levin et al. tested this prediction by mixing evolved cheaters, evolved cooperators, and ancestors in different combinations in the petri dishes, allowing them to form fruiting bodies, and then counting the different spore types in the fruiting body heads. If the evolved cooperators had evolved resistance, they should have a higher representation in the spore head, relative to their un-evolved ancestor.

Indeed, they found this was the case (Levin et al., 2015). While the cheaters always had an edge in getting into the reproductive part of the fruiting body (that's what makes them cheaters!), they had significantly less of an edge against their co-evolving cooperators than against their ancestors. Thus, at least once cooperation has already evolved, there is some scope for individuals to evolve resistance to cheating.

What mechanism maintains cooperation?

Queller and Strassmann had demonstrated that high relatedness maintains cooperation. But what mechanisms are most important for generation such high relatedness? Two key mechanisms might be at play: structured growth and structured dispersal (Inglis et al., 2017).

When the cells within an organism grow with high structure; that is, as they divide, daughter cells remain next to their parents with little movement, this generates high relatedness between interacting cells. Mutants that pop up will find themselves next to other mutants. A selfish cheater will be right next to its daughter, also a selfish cheater, and will lack access to cooperators to exploit. Unstructured growth means a mutant is moving around, likely to bump into a cooperator to exploit. Relatedness will be low. Indeed, this may be why organisms with significant cellular movement set aside their germ-line early in life, to prevent it from being exploited by cheaters (Buss, 1987).

Structured dispersal has a similar effect. If new populations are founded by a small number of individuals, relatedness among neighbours will be high. If dispersal is unstructured, unrelated individuals will all mix together, and relatedness will be low.

To what degree do these different factors impact cooperation? How do they interact? Queller and colleagues tested this by extending their pseudo-organism approach (Inglis et al., 2017). They manipulated growth by either spreading amoebas all across a plate (unstructured, well-mixed growth), or starting a petri dish with a single amoeba and letting it grow outwards (structured growth). They manipulated dispersal by choosing spores to start a new generation from a wide transect of the petri dish (unstructured, well-mixed dispersal) or from a small patch (structured dispersal). This created a total of four possible combinations: structured growth with structured dispersal (highest relatedness), structured growth with unstructured dispersal (intermediate relatedness), unstructured growth with structured dispersal (intermediate relatedness), and finally, unstructured growth with unstructured relatedness (lowest relatedness).

Meanwhile, they started each population as a mixture of mostly cooperators, like those that exist in the wild, and a small amount of cheaters that evolved in the lab. Kin selection theory predicts that under low relatedness, the cheaters might spread, increasing their representation in the population, and under high relatedness, the cheaters would fail to spread.

They passaged the populations through four rounds of this artificial life cycle. Their prediction was that the spread of cheaters would be highest with unstructured growth and dispersal, lowest with structured growth and dispersal, and intermediate with the intermediate combinations.

At the extremes, their predictions were born out, with cheaters failing to spread at all under structured growth and dispersal, and cheaters spreading with unstructured growth and dispersal (Inglis et al., 2017). Interestingly, however, cheaters also failed to spread under unstructured growth with structured dispersal, and cheaters spread the most under structured growth and unstructured dispersal. Inglis et al. explain this as occurring because structured growth, while increasing relatedness, also creates scope for growth mutants to evolve, which exploit the structured growing environment at the expense of cooperation. These results also suggest, interestingly, that structured dispersal might be more important for maintaining cooperation than structured growth.

Structured dispersal prevents cheaters from spreading. An implicit assumption, made not just by this experiment but the vast majority of work on Dicty, is that the act of forming fruiting bodies *is* an act of dispersal. The idea is that by exposing themselves at the tip of the stalk, the spores somehow increase their chance of being dispersed to a new site where there might be food, and that this process increases relatedness.

The Strassmann Queller lab set out to test this assumption, and determine whether or not fruiting bodies actually increasing dispersal (jeff smith, Queller, & Strassmann, 2014). To do this, they explored one specific hypothesis that spores from fruiting bodies are carried on the legs of invertebrates. Using the fruit fly as their test-vehicle, they allowed the insects to fly around an area with fruiting bodies, and then move to clean plates. In some dishes they left the fruiting bodies intact, while in others they disrupted them, breaking the stalks.

The prediction is that, if fruiting bodies aid in dispersal, flies exposed to intact fruiting bodies would carry more spores to the virgin plates, as evidenced by more amoebas growing there. This turned out to be true, demonstrating that indeed fruiting bodies act to increase dispersal, and that invertebrates can serve as vehicles for said dispersal (Jeff Smith et al., 2014).

One aspect not yet fully understood is the role of slug migration in impacting relatedness, conflict, and cooperator in the social amoeba. A number of authors have noted that while the slug migrates, some cells are lost (Jack, Adu-Oppong, Powers, Queller, & Strassmann, 2011; Jack et al., 2015). This might create conflict, as individuals would prefer not to be the ones who die while migrating. Indeed, Queller, Strassmann, and their team have found some evidence of increased genetic diversity (lower relatedness) decreasing distance travelled by a slug, which is possible evidence of conflict. On the other hand, Jack et al. have found that the longer a genetically diverse slug migrates, the less cheating occurs in the fruiting body stage. Exactly why this happens, and how migration impacts conflict, is a question that demands further research.

Do these results extend beyond the lab?

These results are all well and good for telling us what happens in the lab. They show, *in principle*, that high relatedness maintains cooperation, that low relatedness favours cheating and can lead to the breakdown of cooperation, and that structured dispersal is essential for maintaining relatedness and therefore cooperation. They also show, more broadly, the ways in which Dicty and the pseudo-organism lab approach are incredibly useful for testing social evolution theory.

But Strassmann and Queller were interested in whether these results have any bearing on Dicty *outside the lab*. Cheaters can be made to evolve under lab conditions, but do they actually exist in nature? Has there been an evolutionary history of cheaters and cooperators duking it out? Does the conflict created in the lab exist in the forest soils where Dicty grows?

A first step in answering this question is to ask whether high relatedness is the norm in nature. In the lab, Dicty is often grown on smooth petri dishes, a far cry from the complex soil structures which serve as their natural habitat. The Strassmann Queller lab sought to resolve this by creating more realistic growth environments, actually measuring the genetic relatedness of individuals under natural growth conditions, and finally, comparing this to actual results from the wild (j. smith, Strassmann, & Queller, 2016).

Smith et al. found that, even though new populations may be founded by cells only millimeters away from each other, this is sufficient to create high relatedness within fruiting bodies. What looks like a small distance to us, it turns out, is a cavernous gap to amoebas. Further, in comparing these results to densities in natural soil isolates, they found that naturally occurring densities should be sufficient to create high relatedness. Thus, the high relatedness that Strassmann and Queller's team demonstrated to be important for maintaining cooperation in the lab, is generated by natural growth conditions and present in naturally occurring Dicty in the wild (j. smith et al., 2016). It's also worth mentioning that the Strassmann Queller lab has identified some degree of kin discrimination in Dicty, which means that, at least to an extent, they can control whether they form a slug with a relative or a non-relative (Strassmann, 2016). This may also act to increase relatedness in the wild.

These results lend further support to the idea that relatedness matters to Dicty in nature. But what about cheating? A number of the studies mentioned above study the role of cheaters in the social amoeba. But are these an artefact of the lab? Do cheaters exist in nature, and has this played a role in the evolution of *Dictyostelium discoideum*?

Of course, we can't observe the evolutionary history of *Dictyostelium discoideum* directly. But Queller, Strassmann, and their colleagues found a way to explore these questions using genetics (Ostrowski et al., 2015). Evolutionary battles leave a trace in the genome. Different *kinds* of evolutionary battles leaving different, detectable traces in the genome.

Specifically, in this case, the question was, what has the role of cheaters been in the evolutionary past of Dicty? Have cheaters and cooperators both been under positive selection, with each one occasionally gaining an edge until the other regains the advantage? This would be what's called an "evolutionary arms race". Or are they in a "stalemate", each one at a balanced equilibrium selection not acting to increase the frequency of either. Finally, are cheaters simply popping up my mutation, and then slowly being weeded out by selection ("mutation selection balance").

Ostwoski et al. carried out a massive analysis of the Dicty genome, looking for the unique calling card of each of these different types of evolutionary processes. Their results were most consist with a stalemate of sorts, with small fluctuations in frequency occurring because each type (cooperator vs cheater) does better when the other was rare, but otherwise in an equilibrated balance.

This reveals several important findings: first, this type of genome analysis can be used to detect evolutionary histories. Second, cheaters have been a natural part of Dicty's past. And third, it suggests that Dicty's unusual form of multicellularity may have sufficiently reduced conflict to allow cooperation, but without the single-celled zygote stage, there is still enough conflict present to maintain cheating at some frequency, and prevent the full-fledged extreme division of labour we see in most plants and animals.

The social amoeba is also the farming amoeba

As if the social lives of Dicty could not get any more impressive, the Strassmann Queller lab discovered that these seemingly simple microbes possess one other remarkable trait: agriculture (Brock, Douglas, Queller, & Strassmann, 2011). Dicty has a number of relationships with bacteria – it both feeds on bacteria and can be infected by bacteria. But it turns out that dicty also farms bacteria: ingesting them without killing them, carrying them through their dispersal stage, allowing them to colonise the new area, and then harvesting them prudently. Does it get much stranger than that?

Farming is a special form of cooperation, in which one species (the farmed one) is propagated by another, and the other (the farmer) benefits from some services provided by the first. It is remarkable that a simple microbe engages in a behavior that we would otherwise have thought of as relatively complex. However, this raises a number of questions. How do the farmed bacteria and their Dicty hosts find each other? To what degree is the relationship species specific? To what end does the Dicty farm the bacteria?

In a series of studies, Queller, Strassmann, and their students and colleagues set out to unravel the mysteries of Dicty bacteria farming. First and foremost, dicty farm bacteria for the same reason we farm chickens and cows: for food. Dicty eat bacteria, and therefore carry their food with them to new locations. But not all bacteria carried by Dicty are edible. Could they be farming for other reasons, too?

In turns out, yes. First, the Strassmann Queller team looked at whether it's possible that the inedible bacteria farmed by Dicty provide some degree of toxin resistance (Brock, Callison, Strassmann, & Queller, 2016). As the migratory slug

moves, it collects toxins, which it has to discard by shedding cells, which comes with a reduction in fitness. If farming provides toxin resistance, farmers might succeed in not shedding those cells, without suffering negative side effects from the toxins.

To test that hypothesis, one would need to compare fitnesses between both individuals carrying farmed bacteria and individuals without. Dicty, ever the experimentalist's friend, have naturally occurring farmers and non-farmers. Brock et al. tested the prediction with regards to farmers of *Burkholderia* bacteria, found that indeed farmers shed fewer cells during migration without any loss in fitness due to toxins (Brock, Callison, et al., 2016).

In a separate study, the team looked at the toxins produced by the farmed bacteria themselves (DiSalvo, Brock, Smith, Queller, & Strassmann, 2014). For example, Dicty farm *E. coli*, which are inedible to them. The *E. coli* produce toxins which the farmers are resistant to, but which inhibit the growth of non-farming competitor Dicty. Interestingly, the resistance that the farmers have to their own *E. coli* does not lend itself to resistance to other, foreign bacteria (DiSalvo et al., 2014).

Another possible reason for farmer is predator resistance. Dicty are predated by nematode worms. Strassmann, Queller, and one of their students tested whether the farmed bacteria provide defense against these worms (Adu-Oppong, Queller, & Strassmann, 2014). However, they found that not only did the bacteria *not* defend against predators, but that farmers actually did worse than non-farmers in the presence of predators. Thus, farmed bacteria provide food, toxin reduction, and an edge over competitors, but not protection from nematode predators.

Next, the lab turned to the question of how the farmers and the farmed find each other. One possibility is that the bacteria use chemotaxis – attraction to chemicals – to identify and move towards farmers. Shu et al found that not only can the bacteria use chemical cues to find Dicty, but that some of the farmed bacteria are preferentially attracted to Dicty that have *previously acted as farmers* (Shu, Zhang, Queller, & Strassmann, 2018). This suggests an active role for bacteria in the choosing their farmer partners.

Finally, there is the question of who is in charge of the relationship. Do Dicty farmers adjust their behavior, such as how far they migrate or how many bacteria they carry, to provide themselves with agricultural advantages? Or do bacteria manipulate their hosts, controlling whether the Dicty eat the bacteria, when they form spores, etc?

As humans, we tend to think of farmers being in total control. We choose what to grow, when to harvest, how much to eat. But in 2015, The Strassmann Queller lab upended this notion by identifying a causal role for bacteria in the relationship, finding that they could actually initiate farming, and colonise new, virgin Dicty (Disalvo et al., n.d.).

In 2016, they elaborated this study (Brock, Jones, Queller, & Strassmann, 2016). They looked at four farming traits: migration distance (farmers travel less far), symbiont transport (farmers carry bacteria when they disperse), prudent harvesting (farmers don't eat all the bacteria present), and bacterial resistance (farmers don't suffer the negative effects of inedible bacteria that they carry).

In a fascinating result, they found that neither is it the case that Dicty are responsible for all of these traits, nor is it that bacteria are responsible. Instead, short migration distance and symbiont transport are induced by the bacteria, prudent harvesting seems to be influenced but not totally controlled by bacteria, and resistance is a trait of Dicty (Brock, Jones, et al., 2016). In other words, the bacteria are causing the farmers to stop moving so far, and then making the Dicty carry them

to a new location, and then there is some joint control over how many of the bacteria get eaten, and finally, it is the Dicty who are insuring they aren't damaged by their farmed goods. Not only is this a fascinating example of joint agency in a cooperative partnership, but this degree of joint adaptation also suggests that the farming relationship has a long evolutionary history.

Taken more broadly, the vast range of surprising discoveries in Dicty, and the slew of successful experiments carried out by The Strassmann Queller lab, are evidence of a few things: 1) cooperation is important not just in complex higher organisms with brains, memory, and learning, but even in simple microbes, 2) experiments are essential not just for testing predictions from theory, but illuminating new avenues of research, and 3) having a study system, or model organism, that you get to know very well, is an invaluable way to gain a holistic view of various natural phenomena, as opposed to looking at specific traits in isolation.

Cooperation in humans

- **What's special about humans?**
- **Explanations for cooperation in humans**
- **Cooperative games**
- **Inter-personal and cross-cultural differences in cooperation**
- **Psychological systems designed for cooperation**
- **The origins of cooperation in humans**
- **The How: The mechanism of cooperation**

What's special about humans?

And now, humans. Cooperation is, of course, widespread in humans. Parents give their children resources and care, friends spot each other money, and complete strangers provide each other with services and goods. How unusual is this cooperation? Why do we cooperate? Why do some people, or societies, cooperate more than others?

So far we have talked about *general* explanations for cooperation, which work across the tree of life, from viruses to mammals to trees. If these explanations are truly general, then they should work for humans, as well (A. Aktipis, 2016). We are, after all, situated in the tree of life. Then why are we treating them separately, in their own section?

The main reason is that, while evolution *can* explain some human behaviors (this is trivially true), other factors also influence human behaviours (also trivially true). Much of the nature vs nurture, or evolution vs culture, debate can be resolved by recognising that both extremes of the argument are patently true to all serious researchers. Some behaviour can be explained by evolution, but not all of it.

One of the things that makes humans somewhat unique is the degree to which *cultural evolution* shapes our behaviour (K. Laland, 2017). Cultural evolution is similar to genetical evolution in that it shaped by a process of selection, in which cultural replicators, or "memes", that do better at replicating spread through the population (Dawkins, 1976; Dennett, 1995). But there are also important differences, including the ways in which genes and "memes" are inherited, and the much faster rate of cultural evolution. Cultural evolution is significantly less well understood than genetical evolution.

The mechanisms by which cultural evolution works – the way in which, for example, we learn from social cues and adjust our behaviour – are themselves *products of genetical evolution* (Kurzban, Burton-chellew, & West, 2015). This is important to keep in mind. The exact way in which our brain converts social cues into behaviours, ignoring some cues and picking up others, for example, was likely *created by natural selection acting on genes*. The reason one specific method of learning evolved rather than another one was due to its superior gene fitness (or number of offspring associated with possessing it).

However, while genetical evolution makes clear, unambiguous predictions about what traits will evolve (those that maximise inclusive fitness), the same can't be said for cultural evolution. There are two problems here. First, the details of how something is culturally inherited, how learning works, and the nature of the trait influence whether it can and will spread by cultural evolution. Thus, even simple theoretical models, which strip away much of the complexity of the real world, make opposite predictions about whether cultural evolution will increase or decrease levels of cooperation (Kurzban et al., 2015).

But a second problem is that, even insofar as cultural mechanisms are products of genetical evolution, cultural evolution has rapidly pushed humans out of the environment to which they adapted.

A critical assumption of much evolutionary biology is that the environment has not recently changed. If it has, the adaptations the organism possesses will no longer be adaptive. For example, if the arctic suddenly warmed up by 30 degrees, we would have trouble understanding the adaptive significance of a polar bear's fur.

For humans this assumption is almost always violated. For much of our evolution, we lived in one environment, both physically and socially, and within a

blink of an evolutionary eye this has changed. Thus, a cultural mechanism, such as social learning, that *evolved* because it was *adaptive*, may now push cultural evolution in any direction, increasing or decreasing fitness. While we might expect genetical evolution to eventually correct this, there's no reason to expect that has happened *yet*. So, while there is no doubt that the human mind and cooperative behaviour have been shaped by culture (Cronk, 2017; K. Laland, 2017; K. N. Laland, 2017; Nielsen, Mushin, Tomaselli, & Whiten, 2016), there is still significant theoretical work to be done, to say nothing of empirical work, before we understand the role of culture in shaping human cooperation.

What can we say about human cooperation then? As before, we are interested in "evolutionary explanations" for cooperation. Remember, our definition of cooperation is: "a behaviour which provides a benefit to another individual (recipient), and which is selected for because of its beneficial effect on the recipient" (West et al., 2007).

A basic premise of much of the work on human cooperation is that complex psychological mechanisms that govern our actions are products of evolution. This is a relatively uncontroversial idea now (Cosmides & Tooby, 2013). This means, as with before, we are not talking about behaviours which are designed to serve some other purpose, but incidentally give some benefits to others. All actions have myriad effects on many individuals. For example, if you toss your scraps out the window of a 5 story apartment because you don't want them there, and this inadvertently feeds someone's pet pig on the ground floor, this would benefit others. But the puzzle is why would we have psychological mechanisms specifically designed to benefit others, not why someone would chuck scraps out their window.

For that reason, it's also important to keep a distinction between proximate and ultimate explanations, as we did with other organisms. For example, a proximate explanation for why someone eats food is that their glucose levels have dropped below a certain threshold, triggering a reaction in the brain, etc. But the ultimate explanation is that food consumption increases fitness, and selection constructed that brain pathway *for the purpose of having more offspring.*

Thus, "humans cooperate because it feels nice" is not a sufficient explanation. Humans cooperate because they have morals is not sufficient either. Even, humans cooperate because the brain rewards them with chemicals is not sufficient. Our question is: why does it feel nice? Why do we have morals? Why does the brain reward us for cooperation and not selfishness? What selection pressured shaped these psychological structures?

Explanations for cooperation in humans

So far, we have focused on cooperation for *indirect benefits*, which is when benefits accrue to relatives, and is explained by the theory of kin selection. Can cooperation in humans be explained by kin selection, too?

The short answer is, yes, some of it. The most obvious example is parental care, in which benefits are targeted to offspring, who share half of a parent's genes. Beyond that, there is a wealth of evidence that humans preferentially help relatives, that humans help closer relatives more than distant relatives, and that humans evolved in small groups where relatedness might have been high (Kurzban et al., 2015).

That said, we have already covered kin-selected cooperation in detail. Here, we turn to focus on cooperation for *direct benefits*, in which the benefits of cooperation return (eventually) to the co-operator. There are four common explanations for direct-benefit cooperation (Figure 3 is repeated here for convenience), which I'll summarise briefly here, but discuss more in depth later:
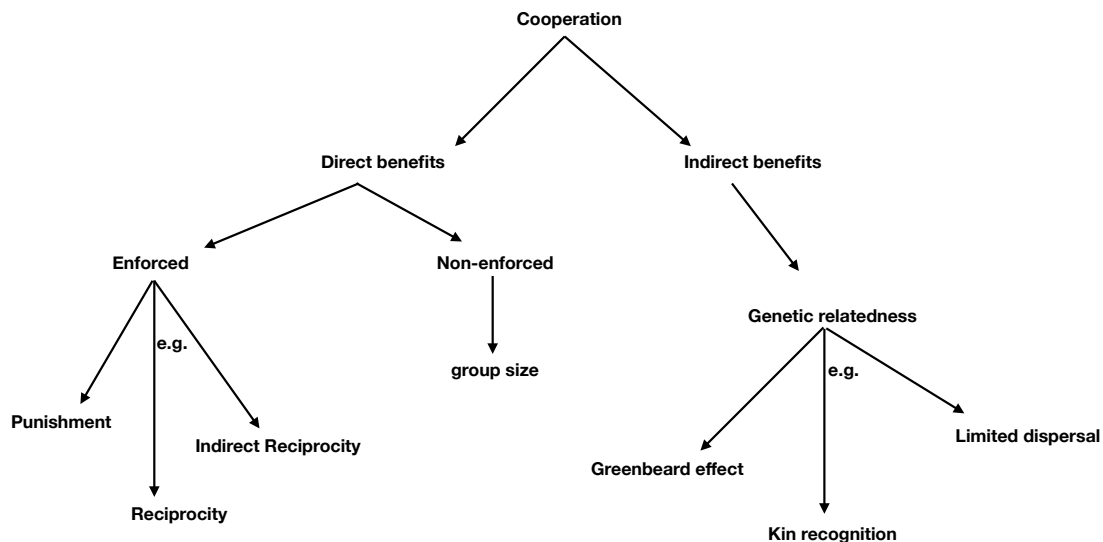
1. Reciprocity

Trivers (1971) pointed out that individuals might cooperate if there is a probabilistic likelihood of that cooperation being returned to them. For example, I might lend you money if you are reliably going to lend me money in the future. More precisely, the *expected* return of benefits should outweigh the costs of cooperating.

2. Indirect reciprocity

Individuals might cooperate even if the recipient doesn't return the favour, if cooperating is correlated with receiving benefits from some third party. For example, if cooperating gains you some reputation points, which causes you, later, to be given benefits from some external observer, and these benefits outweigh the initial costs, that would be sufficient reason to cooperate.

3. Punishment (avoiding costs)

It has been argued that one might cooperate in order to avoid the costs of being punished. This explanation is still controversial, both in experiments and in theory (Gardner, West, & Barton, 2007; Hauser, Nowak, & Rand, 2014; Krasnow, Delton, Cosmides, & Tooby, 2015; Kurzban et al., 2015). On the one hand, some work has suggested that punishment can favour *any trait*, including cooperation. However,

other work has suggested that punishment never favours cooperation when it wouldn't otherwise be favourable. This is still an active area of research and an unanswered question,

4. Coordination (morals)

There is strong evidence that humans are cross-culturally moralistic (Kurzban et al., 2015), judging actions as right or wrong and seeking justice for wrongs. A simple explanation would be that morals promote pro-social behaviour. However, morals often judge behaviours which do not harm the group, or are even positive towards the group, as morally wrong. Further, human morality is often non-consequentialist, judging actions not just on their consequences, which doesn't fit with simply improving group welfare. An alternative explanation is that morals are used to choose sides in conflicts, coordinating individuals into large, cooperative groups. Doing so creates power asymmetries, allowing them to settle disputes easily at low cost (Kurzban et al., 2015). This "bandwagon" interpretation of morals would explain why they shift over time so much – they are used less for upholding specific behaviours, and more for coordination.

These are explanations of cooperation *in principle*. However, individual behaviours are often complex, and carried out under varied circumstances with many different types of interactants. Pinpointing a specific explanation is usually complicated. If the explanation is reciprocity, why does the player probabilistically expect a return? How many observers need to witness a cooperative act for indirect reciprocity to work? How is institutional punishment different from individual punishment? When does the benefit of coordinating through moral systems outweigh the cost?

There are two approaches to answering these questions: theory and data. Theoretical work involves constructing mathematical models. Often these models utilise something called game theory, which treats individuals as players (e.g. person 1 and 2) which can adopt strategies (e.g. cooperate or defect). Players get payoffs (e.g. money) depending on their own strategy and the strategy of the other player. The mathematics of game theory tell us which strategy all players will adopt if they are trying to maximise their payoffs.

The most basic form of game theory model assumes that the players (humans, in the case), are rational actors trying maximize some utility function. For example, they might be trying to maximise their personal wealth.

Experiments can then test these predictions by having actual individuals play the games, either in person or on a computer. Participants may be given real money, and allowed to play the game, and the strategies they adopt can be observed or measured. Do they maximise their wealth? Or do they give away more money than would be expected if they were maximizing their own wealth alone? Are they *cooperative?*

Field work can also look how people behave in real life, for example by looking at what they do in game shows, or measuring church donations, etc. Finally, comparative work can look across cultures or institutions to see how they differ in cooperative strategies.

However, all of these deserve a few grains of salt before proceeding.

First, as with any models, the assumptions made by game theory are often violated. This doesn't make the models wrong or useless, but it's always worth keeping in mind what assumptions have been made, how realistic they are, and how important violations to the assumptions might be in practice.

Second, experiments with humans are difficult. Just as one example, a number of studies have found that humans, playing games, are more prosocial than we would expect if they were merely selfish beings. For example, they give away more money than they should. However, a recent study (Burton-Chellew, El Mouden, & West, 2016) has suggested that participants may often simply *not understand the game*. In other words, the games in these experiments are sometimes fairly complicated. It *may* be that the person is trying to maximise personal wealth, but just not doing a very good job at it. Again, this doesn't mean all human games experiments are wrong, but it's worth keeping an eye out for whether there is convincing evidence that the players understand the game.

Third, cross cultural experiments have to handle a huge number of confounding variables. This isn't impossible, but requires great care, both in study design and in statistics. Scientists should go to great lengths to convey that this care has been taken (Kurzban et al., 2015).

Finally, before we dive into it, a note on human cooperation: how cooperative are we? We asked ourselves this question at the very beginning. Indeed, our starting point was humans may take cooperation for granted, because it is so widespread in our society. It is often assumed, and sometimes explicitly stated in the literature, that humans are somehow *especially* or *uniquely* cooperative, demanding some *extra* explanation, beyond that in animals.

But as we have seen, humans are significantly *less* cooperative than some organisms. Take our friend *Dictyostelium discoideum*, or eusocial insect colonies, in which a large proportion of organisms *give up reproduction altogether* to help other individuals. We certainly don't reach those levels of cooperation as humans.

That's one of the great benefits of having a theory – the theory of evolution by natural selection – that unites an entire discipline. It allows us to put things in a framework, comparing and contrasting across the tree of life.

What's special about human cooperation isn't its extremity. Instead, it's the complexity introduced by things like memory, learning, culture, emotion, institutions. That's something worth diving into.

Cooperative games

In theory

As mentioned, one approach to understanding cooperation in humans is game theory. Game theory is a mathematical approach to modelling, and, more broadly, a way of understanding behaviour. A game theory model consists of players, in this case it's usually human individuals, strategies, which are how the humans can behave, and payoffs, determined by the strategies of the players. The classic example is the prisoner's dilemma game.

In this scenario, there are two prisoners (players A and B) which have been arrested for a crime, and are each offered a chance to pin the crime on the other prisoner (they are in solitary lock up, and can't communicate with each other). We call pinning the crime on the other prisoner *defecting* (strategy 1) because they are betraying their partner in crime. We call remaining silent and saying nothing

cooperating (strategy 2), because they are helping their partner in crime. Depending on their own strategy, and the strategy of their partner, they get different payoffs.

If players A and B both defect, betraying each other, they each get two years in prison.

If player A defects (tattling on the other) and player B cooperates (remaining silent), A serves no time in prison and B serves three years (and vice versa).

If both players cooperate, remaining silent, they both serve one year in prison.

	B Cooperates	B Defects
A Cooperates	1 year 1 year	0 years 3 years
A Defects	3 years 0 years	2 years 2 years

Figure X. Prisoner's Dilemma. The cells contain payoffs to players A and B depending on both A and B's strategies (i.e. cooperate or defect). Shaded portions of a cell show payoffs to player A and white portions show payoffs to player B. The strategy players will settle on is the one for which neither player can change strategies to get a higher payoff. In this case, it's Defection for both A and B (Defect/Defect).

Thus, the players are prisoners, the strategies are to cooperate or defect, and the payoffs are jail time. What will the players do? That depends on what they value, or what *function of payoffs they are trying it maximise* (which is another way of saying *what they value*). If they are purely self-interested, and trying to maximise

time not spent in jail, and they are rational actors (make consistent choices based on what they value), they will always defect.

Why? Look at figureX. Let's say both players are cooperating, and going to get 1 year in prison. If I am player A, switching to defect, I go from 1 year in prison to 0 years, so that would be the wise choice. The same holds for player B. Now, we are both defecting, and going to get two years in prison. If I am player A, considering switching to cooperate, I would be going from two years in prison to three years, which is worse off. The same holds for player B. Thus, neither of us will switch. Defect/Defect is an *equilibrium,* which both players will settle on if they are rational actors behaving selfishly.

Why then, would individuals ever cooperate? The above is a mathematical model, and can't be wrong, given the assumptions. So, it can only be wrong if one of the assumptions is wrong. For example:

1. The actors might not be rational. They make mistakes, have imperfect information, etc.
2. There may be other strategies we didn't include – some third alternative available to them, such as a conditional strategy (cooperate if X, defect if Y)
3. They may be playing a different game, such as a donation game, with different strategies and payoffs. Or the game may be repeated, and individuals may remember previous games, and condition their behaviour on the outcome of those games.
4. They may be maximising something other than self-interest. This is akin to discovering that organisms don't maximise personal fitness, but inclusive fitness.

These are just a few examples. Indeed, an entire field has blossomed around relaxing these assumptions and others (B. Allen & Nowak, 2013), and looking at the consequences. And few people have done more for that cause in the last couple of decades than Martin Nowak of Harvard, who has elaborated the prisoner's dilemma and other games in nearly every way imaginable, with fascinating results.

Zero-determinant strategies

In the simple prisoner's dilemma game described above, defection always wins. This is bad news for the evolution of cooperation. However, if the game becomes repeated, or iterated, there becomes scope for reciprocity, in which co-operators benefit from recouping the costs via a cooperative partner.

Nowak and his colleagues, including Arne Traulsen of Max Planck, became interested in a particular type of strategy in this iterated games called "zero-determinant strategies". With these strategies, a player can enforce a linear relationship between their own payoffs and the payoffs of a co-player. An example of this would be a "tit-for-tat" strategy, in which a player cooperates in the first round, and then, in subsequent rounds, does whatever her partner did in the previous round. It turns out that for this, and other zero-determinant strategies, it can be shown mathematically that the player adopting the zero-determinant strategy enforces a linear relationship between their payoffs, preventing that player from being exploited.

But it gets even more interesting. A subset of these strategies engage in "extortion", not only pinning their partner's payoffs to their own, but actually

ensuring their partner gets a lower payoff then themselves. The problem is that, as everyone in the population begins to adopt an extortion strategy, there is no one left to extort, and the strategy is invaded by co-operators. Nowak and his team set out to find out whether this was always the case (C. Hilbe, Nowak, & Sigmund, 2013)

Using game theoretic models, Nowak and his team found that indeed, in large populations, extortion has only a fleeting existence, before it is replaced by cooperation. However, if the two players are from separate populations, evolving at different timescales, the situation changes. This might occur, for example, if one player is from a small, fast breeding species, and the other is from a larger, slower breeding species. In this case, extortion can spread in the more slowly evolving species, to the point where that species "enslaves" the other with extortion (C. Hilbe et al., 2013).

Nowak and his colleagues also directly compared extortion with "compliance", which is when the player ensures his own payoffs don't exceed his partners. To do this, they used a mathematical approach called "adaptive dynamics", which allows one to look at the invasion of rare mutants into a population. They found, interestingly, while extortion gives way to cooperation, compliance ultimately leads to a population of selfish individuals (Christian Hilbe, Nowak, & Traulsen, 2013). This reveals the counterintuitive nature of games – it's often not the "best" strategy (from our moralistic perspective) that leads to the "best" outcome for individuals or populations.

Hilbe et al. extended this result to consider the effect of variation in group sizes. Group size influences relatedness and the likelihood of reencountering the same partner (Hamilton 1964, Frank 1998, Krasnow, Delton, Tooby, & Cosmides, 2013), and therefore might change how different strategies fair. Hilbe et al. found that, while at large group sizes compliance (and other "generous" strategies) ultimately pave the way to selfishness, at small group sizes such generous strategies can maintain cooperation (Christian Hilbe, Wu, Traulsen, & Nowak, 2015). The question remains how cooperation would be maintained at large group sizes.

Nowak and his colleagues extended his results to repeated games with any number of players, providing a formal theory for zero-determinant strategies (Christian Hilbe, Wu, Traulsen, & Nowak, 2014). They also showed that coordination between players can resolve some of the problems posed by large groups, and, more broadly, that zero-determinant strategies are evidence that individual players can exert control beyond their own outcomes.

The evolution of hypocrites

Zero-determinant strategies are about reciprocity – ensuring that the benefits you receive from other co-operators are at least as good if not better than those your partner receives. But direct returns from a partner aren't the only way to benefit from cooperation. The other possibility is that you receive benefits from a third party, or via *indirect reciprocity*. A classic example would be reputation effects: when individuals gain a reputation based on their actions, and cooperative decisions are made based on those reputations, it can pay to cooperate and therefore gain a good reputation.

However, Nowak, Ohtsuki and Iwasa were interested in a critical assumption that most models of indirect reciprocity make: that all interactions are observable. In real life, some interactions may be observed by others ("public"), while others may

be viewed only by your interacting partner ("private"). If both public and private interactions occur, it might pay to cooperate in public, when one is being observe, but defect in private, when one's reputation won't be affected (Ohtsuki, Iwasa, & Nowak, 2015).

Ohtsuki et al. decided to model cooperation when some actions happen in private and others in public, and players can adjust their strategy accordingly. Following standard approaches to indirect reciprocity, players can cooperate with "deserving" players (those with "good" reputations for being cooperative) and defect against bad players. They found that under some circumstances an "honest" strategy evolves, in which players cooperate with deserving players both in public, when being observed, *and* in private. However, in other circumstances a hypocritical strategy evolves, in which players cooperate in public but defect in private. Critically, "honest" strategies are only evolutionarily stable when the chance of being observed is high. In other words, as interactions become increasingly private, hypocrisy prevails (Ohtsuki et al., 2015).

This suggests that hypocrisy poses a threat to cooperation. This *might* suggest that humans would have mechanisms to punish or prevent hypocrisy as a strategy. Interestingly, Nowak et al.'s colleagues at Yale used experiments to probe the psychology of hypocrisy, and found that this might be true (J. J. Jordan, Sommers, Bloom, & Rand, 2017). Jordan et al. showed participants vignettes in which people engaged in behaviours such as honesty, hypocrisy, etc. They found that verbal condemnation of others' behaviour signals "goodness" even more than simply stating one's moral beliefs. Further, people judge hypocrites, those who condemn a behaviour in others they themselves engage in, even *worse* than directly lying about their own beliefs. Finally, "honest" hypocrites can escape this negative judgment by admitting to transgressing at the very act they've condemned (J. J. Jordan et al., 2017). These strong and somewhat counterintuitive responses to hypocrisy suggest that the game theory models of Nowak et al. may have some bearing on how our psychological systems evolved to handle hypocrisy.

And lest we think observability only matters in mathematical models and the lab, Nowak and colleagues have tested this idea in a large scale field experiment (Yoeli, Hoffman, Rand, & Nowak, 2013). They partnered with a utility company, and monitored a program in which people could cooperatively participate in reducing blackouts. This is a public good, because people had to pay a private cost to contribute to something that benefited the group, or neighbourhood. They found that observability of participating in the public good tripled participation, a stronger effect than even paying people to participate.

Relaxing assumptions: Motives and intentions

So far we have talked about altering the strategies of players and the types of interactions that occur. Now we turn to a series of ways in which standard assumptions in these models can be relaxed, and how this impacts the evolution of cooperation.

Standard models assume players have no information about others decision making process, or at least ignore "motives" in decision making, and focus only on actions. Thus, players make decisions based on what other players *do*, not *why* they do it. In real life, however, we seem to focus heavily on motivations, sometimes even more than actions or outcomes. Why?

Nowak and colleagues set to test this out by elaborating standard interaction games to allow individuals to observe other players' behaviour which might reveal their "motivations". Specifically, they allowed players to look at the potential payoffs of cooperating before acting, or "cooperate without looking". Then, the second player, observing whether they looked or didn't, can choose whether to continue the interaction. In a series of two papers, Nowak and his team showed that "cooperate without looking" could, under some circumstances, evolve (Christian Hilbe, Hoffman, & Nowak, 2015; Hoffman, Yoeli, & Nowak, 2015). They interpret this is as occurring because a player who looks might be tempted to cheat by large incentives, and therefore a non-looking partner is more reliable. Further, they suggest this may go some way towards explaining why individuals value friends who don't keep track of favours, morally principled people who decide their actions before checking the outcomes, and even love itself: a commitment that holds regardless of circumstances ("love without looking").

In a relatedness study, Nowak's colleagues at Yale looked at the role of intentions in actual experiments (David G. Rand, Fudenberg, & Dreber, 2015). They allowed players to have "intentions", but then make mistakes when implementing actions. In other words, their intentions might not line up with what they do. Further, they either allowed intentions to be observable or hidden. They found that when intentions are observable it favours cooperation, and that players end up conditioning their behaviour on intentions not actions. This adds further evidence that internal processes matter for the evolution of cooperation.

Relaxing assumptions: Information about partner and self

Another common assumption of game theory models is that prior to a first interaction individuals have no information about their partner, and that, either all games are sequential, or simultaneous games do not change an individual's information on which she can act.

In a series of models, Nowak and his colleagues relaxed both of these assumptions (Manapat, Nowak, & Rand, 2013; Reiter, Hilbe, Rand, Chatterjee, & Nowak, 2018). First, they were interested in trust. Standard models predict that in a first interaction individuals should never trust their partner. However, in experiments that doesn't seem to be the case. They believed this was because individuals may have some information about partners before they ever interact with them, for example through reputation. Relaxing the assumption of no prior information, they found that trust could evolve (Manapat et al., 2013).

Second, they considered the introduction of "crosstalk" to standard models. In real life, we engage in multiple interactions simultaneously, and one interaction may influence our behaviour in another. They extended traditional models to allow this kind of "crosstalk" between interactions, and found that doing so impedes the evolution of cooperation (Reiter et al., 2018). This raises the question of what might further protect cooperation when multiple simultaneous interactions occur, and these interactions can influence each other.

Relaxing assumptions: Optionality

A basic set up of most game theory models is that there are players and strategies, these players interact with each other and play their strategies. Normally, individuals don't have control over who their partners are or whether or not they interact. In

reality however, humans often choose whether to engage in interactions: if a store offers them a bad deal, they'll walk away.

What happens to cooperation if we incorporate optionality? Nowak and his team looked at this problem by developing indirect reciprocity models in which co-operators walk away from interactions with known defectors (Ghang & Nowak, 2015; J. Olejarz, Ghang, & Nowak, 2015). In addition, they adopted their own approach of allowing some information to be public and some private. They found that in large populations cooperation could evolve, given that information was sufficiently public, the average number of rounds per person was sufficiently high, and the benefit to cost ratio of cooperation was sufficiently large.

Relaxing assumptions: Player ability

Finally, Nowak and colleagues have looked at what happens when humans aren't very capable. A baseline assumption of game theory models is that players have perfect knowledge of other players' strategies and payoffs. However, humans are messy and imperfect. What happens when we incorporate this into models?

In a standard ultimatum game, one player proposes a division of a sum of money between himself and his partner. For example, if the first player is given 10 dollars, they can take 5 and give the other 5 to their partner, or take 9 dollars 99 cents and give 1 cent to their partner. The other player can choose whether to accept the division. If players are perfectly rational and self-interested, they should accept any nonzero offer. However, in experiments players often reject non-fair divisions.

Nowak et al hypothesised that this might be explained by individuals being bad at judging payoffs and strategies, and they incorporated this into a model. They found that such mistakes favour fairness, because it leads to a relative advantage in repeated interactions, at least when not all strategies are always present in the population (D. G. Rand, Tarnita, Ohtsuki, & Nowak, 2013).

Inter-personal and cross-cultural differences in cooperation

Above we looked at the kinds of cooperative strategies that can evolve in theory, some of which has been supported by experiments. This kind of game theoretic work tells us what strategies evolve under what conditions, and they can also be tested by having real humans play games in the lab. Occasionally this also extends to the field, when real life scenarios closely mimic games (as in the example of the utility company and the black-out preventing program). More broadly, this type of work is used to revealing underlying principles.

But another question is: what explains real life differences in cooperative behaviour, either between individuals or across cultures?

We've already mentioned, for example, how relatedness might drive these differences. Higher relatedness promotes higher cooperation. Humans may no longer live in small, disparate groups like we once did. But a similar effect can be generated by "static" network structures, which increase the likelihood that we

interact with the same players repeatedly (David G. Rand, Nowak, Fowler, & Christakis, 2014). But what other aspects of culture, society, or personality might explain variations in cooperation?

Religion

One contentious example is religion. An obvious candidate role of religion is to promote prosocial behaviour, but the evidence for this has been controversial and inconclusive. In a series of studies using both field and lab work, Nowak and colleagues tested the hypothesis that religion promotes cooperation (Everett, Haque, & Rand, 2016; David G. Rand, Dreber, et al., 2014).

First they conducted a correlational study in a Christian church. Participants (who had just attended a church service) were given a Christian passage about charitable giving to read. Afterwards, they were asked to rate, on a scale of 1-7, how much they connected to the passage. Finally, they played a Prisoner's Dilemma game. Rand et al. found a correlation between how much an individual connected to the religious passage and whether they cooperated in the prisoner's dilemma. They interpreted this to mean that religion may act as a prime for prosociality.

Second, they carried out a similar study online, allowing individuals to read religious passages and then play a prisoner's dilemma game. Here, they found that while *Christian* individuals were more cooperative after reading *Christian* passages, similar Hindu and secular passages had no effect on Christians, and that *none* of the passages had an effect on non-Christians (David G. Rand, Dreber, et al., 2014). Together, these studies suggest that while religion may act to increase pro-sociality, the selectivity of it is nuanced and not yet fully understood.

In a follow up study, Nowak and colleagues looked at the role of religion in making donations to others (Everett et al., 2016). Participants played an online game in which they had the opportunity to donate some portion of money given to them to another player whose religion was displayed during the game. Everett et al. found that those players who stated that religion was very central to their lives were more likely to donate, regardless of whether recipient was religious or atheistic. This lends further support to the idea that religion may act as a prime for pro-social behaviour.

Nationality

Religion offers one explanation of between-group differences. What about nationality? Nowak and colleagues explored this problem by comparing cooperative behaviour between players from the USA and players from Romania (Ellingsen, Herrmann, Nowak, Rand, & Tarnita, 2012). Players engaged in a series of games, starting with a simple prisoner's dilemma game and becoming more complex. While they found that there *were* cross-cultural differences, these differences were complicated.

In simple games, they found no difference between USA and Romanian groups. In more complicated games, they found that USA groups were more cooperative, which they attributed to USA players' willingness to punish cheaters. In order to be fully convinced by these results, one would want to be sure both groups understood the game equally (Burton-Chellew et al., 2016). Regardless, this study shows how one might go about looking at cross cultural differences in cooperation, and suggests these differences may be real (Ellingsen et al., 2012).

Inequality

People prefer an equal distribution of resources (Nishi, Shirado, Rand, & Christakis, 2015), and yet often resources aren't distributed equally. What effect does this have on cooperation, and to what degree does inequality explain differences in cooperation?

Nishi et al. conducted a large scale experiment in which players could trade and share money over a series of games. A key variable which they found to matter was the degree of wealth visibility. When wealth was visible (as opposed to invisible), inequality became greater, and cooperation decreased (Nishi et al., 2015). This suggests that a combination of wealth visibility and inequality (the former of which may promote the latter) may disfavour cooperation. How exactly this translates to real societies isn't clear, but begs the need for large scale comparative work on the question.

Interpersonal differences

Average cooperative behaviour varies between groups. But do individuals also vary? Of course individuals vary from moment to moment, or interaction to interaction. But are some individuals fundamentally more cooperative than others?

Nowak and colleagues analysed data from over a thousand individuals across thousands of cooperative games. They found that individuals *are* consistent in their cooperative behave, both between different games, and between games and real life settings. They found that an individual's behaviour even correlates between interactions over a hundred days apart (Peysakhovich, Nowak, & Rand, 2014). This suggests that individuals may posses a "cooperative phenotype", which determines some consistent level of cooperation, and varies between individuals. If true, this opens up a fascinating series of questions: what explains this difference? Can we pinpoint a genetic basis? Equally exciting, this result also suggests that results from games may have further reaching implications than previously thought, as it would mean cooperative games are a good measure of an individual's fundamental *cooperativeness* (Peysakhovich et al., 2014) (of course, given the experiment is well designed!).

Nowak's colleagues lent further support to this idea in 2016, by showing that three "cooperative phenotypes" were sufficient to provide high predictive power in explaining individuals' cooperative tendencies across games (Epstein, Peysakhovich, & Rand, 2016). They dubbed these three phenotypes the "good" (individuals who usually cooperate), the "bad" (who rarely cooperate), and the "unflinchingly selfish" (who never cooperate). This does not necessarily mean that there are only three cooperative phenotypes in reality. But the fact that strong predictive power can be gained with non-continuous cooperative traits suggests that there *may* be discrete cooperative types amongst humans, and at the very least lends credence to the idea that cooperation is a self-consistent trait. Importantly, these results suggests that it's not just "tendency to cooperate" that is consistent, but "how to condition one's own behaviour on others" (Epstein et al., 2016).

Cooperation interventions

So far we have discussed what conditions can explain the emergence of cooperation in a population, and sources of interpersonal and intercultural variations in cooperative behaviour. What about interventions to *change* cooperative behaviour?

David Rand of Yale has studied this phenomenon in depth (Kraft-Todd, Yoeli, Bhanot, & Rand, 2015). For example, in one study they considered the role of norms in shaping behaviour. A well-known phenomenon is that humans copy successful behaviour. If an institution creates an environment in which "cooperation" is a successful behaviour (e.g. by punishing defectors), will this carry over into other settings?

Kraf-Todd et al.'s study suggests that the answer is yes. They had participants play a cooperative game in an environment in which cooperation was either a successful behaviour or not, and then observed individual's behaviour in a subsequent, one-shot game. They found that when cooperation was previously a successful strategy (a positive "norm"), individuals were more prosocial in the new setting. Of course these two "environments" are not as varied as those one might expect in real life, but the result suggested a role for institutionally created cooperative norms.

Rand and colleagues looked at this more explicitly in a further study, which analysed the role of top-down institutional enforcement of cooperation (Stagnaro, Arechar, & Rand, 2017). They compared the performance of players who had experienced different self-reported qualities of institutionally enforced cooperation (e.g. police) in different cooperative games. They found that players that reported experiencing better cooperation enforcement from institutions were more prosocial in games. Of course, this is only a correlational result. However, they also experimentally manipulated institutional punishment in a series of games, and subsequently observed the same result: individuals who had experienced stronger cooperation-enforcing institutions subsequently were more prosocial. Interestingly, the quality of institution experienced had no effect on individuals' subsequent *punishment* behaviour per se, simply their willingness to give to others. A weakness of this latter approach is that it could be explained by simple conditioning. But Stagnaro et al. argue that, together, their results show convergent evidence for the role of institutional cooperation enforcement in shaping individual behaviour outside institutional settings (Stagnaro et al., 2017).

Rand and colleagues have used theory to consider the roles of legal systems in maintaining cooperation. Laws are one way to punish defectors and favour cooperation. But mathematical models looking at the "coevolution" of laws and cooperation suggest that the speed at which laws are changed might matter (Roithmayr, Isakov, & Rand, 2014). Their model shows that, under some conditions, if the law updates too quickly, citizens don't have time to adapt their behaviour accordingly, and cooperation isn't rewarded. However, if laws update too slowly, new defector strategies can't be punished, and cooperation can collapse. This suggests that at least in principle there is a critical rate of law adjustment optimal for maintaining cooperation (Roithmayr et al., 2014).

More broadly, Rand and colleagues argue that the evidence points towards interventions based on observability and norm-enforcement working better to promote cooperation than those based on altering the costs and benefits of cooperation (e.g. by directly paying people to cooperate) (Kraft-Todd et al., 2015).

However, all of the above examples focus on interventions to improve cooperation with neighbours and groupmates. But one of the great cooperative challenges facing society today involves cooperating not with your peers, but with

future individuals, who may not yet exist. Suffering caused by global warming will be most strongly felt in the future. Reducing carbon use today is costly now, and the benefits accrue to others in the future. Can this type of cooperation ever be favourable?

Nowak, Rand, and colleagues devised an ingenious "intergenerational goods" game to answer this question. Individuals have an opportunity to use up a valuable resource, which will affect players in subsequent rounds. They found that when individuals make decisions independently, the resource is almost always overexploited. However, when decisions are made by binding votes, "cooperating with the future" can be favoured. They argue that this is because democratic voting prevents a majority of co-operators from being exploited by a small number of defectors, and because it makes co-operators feel as though their cooperation isn't futile (Hauser, Rand, Peysakhovich, & Nowak, 2014). If this result extends beyond the lab-game scenario, it will have serious and exciting implications for real life policies and interventions.

Psychological systems designed for cooperation

We have focused on the evolutionary *reasons* for cooperation in humans. More precisely, on the conditions that favour cooperation, either over evolutionary time or within a series of interactions. The approach to studying this aspect of cooperation is to model the conditions under which cooperation should evolve, and see whether that matches human behaviour, either in real-time games or across cultures.

But if we take as given, for a moment, that cooperation *has* evolved in humans, what kinds of predictions does this make? One obvious one is that if cooperation is a human adaptation, the brain should have evolved, complex psychological systems for regulating and controlling cooperation. In other words, if we cooperate for an evolutionary *reason*, then this should be evidenced by somewhat fine-tuned cooperative behaviour, decision making, and reasoning.

John Tooby and his colleagues at UC Santa Barbara have been looking at that question for some decades now, and are leaders in the field of cooperative evolutionary psychology (Cosmides & Tooby, 2013). They are interested in the question of what kinds of psychological mechanisms have evolved to control cooperation.

How to choose a cooperative partner

Here's an example. Many animals, including humans, have to forage for resources, and therefore the brain should (and does) possess a number of tools to regulate this behaviour. For example, we should have mechanisms in place to judge the difference between good and bad resources, and to ensure that the costs we expend foraging (in terms of energy, risk, etc.) are outweighed by the benefits of the resources we acquire.

But humans also forage cooperatively, in groups, and thus we should have evolved mechanisms specific to social foraging. Research on real modern hunter

gatherers, such as Hadza, shows that humans forage in groups (Berbesque, Wood, Crittenden, Mabulla, & Marlowe, 2016). Further, it shows that different foragers gather different amounts, that foragers eat some of the food they gather while foraging, and that this varies between individuals (Berbesque et al., 2016). Thus, there will be different quality partners, and humans should have mechanisms for distinguishing between good and bad partners. Do such mechanisms exist?

In 2012, Andrew Delton and Theresa Robertson designed a study to find out (Delton & Robertson, 2012). One key difference between asocial and social foraging, they argued, is that with social foraging, immediate costs and benefits are all that matter. When judging a potential partner, if they mess up, and drop some berries, the benefit they offer is lower. However, with social foraging, what matters more than immediate metrics are a foragers long term *intentions*. Someone who drops berries while taking a great risk to wade into a dangerous swamp for the group might be more valuable than someone who accidentally picks up a few berries while avoiding risks.

Delton and Robertson showed a number of participants a series of vignettes in which individuals, marooned on an island, engaged in different foraging behaviours like those described above. They then conducted categorization tests to answer two questions: 1) do individuals form categories that distinguish between foragers who pay small and large costs 2) do people value benefits accrued when these aren't intentional? If social foraging has shaped the brain, they people should form categories around costs incurred, regardless of benefits (when these costs are in course of helping the group) and benefits should only matter if they are intentional.

How does a researcher determine what categories an individual forms? Tooby, Delton, and their colleagues have devised an ingenious method, adopting an old technique for studying racial and gender categories for the purpose of studying cooperation. The idea is simple: humans remember between-category differences more strongly for within-category differences. If shown a series of plants and animals and their Latin names, and later asked to match names to pictures, one is more likely to mismatch one plant name to another plant then a plant name to an animal, because these are real categories. Similarly, in attributing foraging actions to individuals from the vignettes, the matching mistakes reveal what categories an individual truly stores.

Delton and Robertson found both of their predictions to hold, showing that individuals categorise others based on the costs they incur for the group, and ignore incidental benefits, instead categorising based on intentions to help the group. (Delton & Robertson, 2012). Together, this suggests the brain has been shaped for purpose of social foraging.

Tooby and colleagues used this same approach to study how the brain judges free-riders (Delton, Cosmides, Guemo, Robertson, & Tooby, 2012). Any cooperative task is vulnerable to free-riders which benefit from the group resource without incurring the appropriate costs to contribute to it. If cooperative behaviour has been important for humans, then we should have mechanisms for judging free-riders. Importantly, we must distinguish between those who don't contribute by accident and those who intentionally behave selfishly.

Tooby and colleagues first set out to test whether subjects judge people as free-riders. They presented subjects with scenarios of individuals foraging for food. Occasionally a number of these individuals came back with less food than others, but there was an important distinction: some ate the food they found, while others simply lost it. Their prediction was that if free riding matters, the subjects should

judge those that ate the food more harshly, even though the absolute returns of food were the same.

Delton et al. found that this was the case. To further ensure this wasn't a fluke of the specific scenarios described, they did a similar experiment in which, rather than having some of the foragers eat their food, they had some simply expend less effort. Again, they found that this mattered to subjects even when it didn't affect the total amount of food returned.

Together these results suggest that subjects categorise individuals as free-riders, and that this is an important concept to them. But Tooby and colleagues wanted to know whether free-riders was a unique concept, designed specifically for the context of behaviours like cooperative foraging, or if subjects were simply identifying a more general category of "moral violators" (in other words, "bad people").

To do this, Delton et al. added to the vignettes of free riders examples of other, non-freeriding, immoral behaviour, such as fighting. They found that, rather than lumping all bad behaviour together, subjects maintained a meaningful categorical distinction between free-riders and other, non-free-riding moral violators (Delton et al., 2012). Taken all together these studies suggest that free-rider detection is a real psychological mechanism, shaped by evolution (Delton & Sell, 2014).

What does someone have to do to be deemed a free-rider? One option is to wait to see if someone contributes less but still partakes in the pie. Alternatively, we might be able to prevent such costs to the group if we judge people who simply don't partake in the collective action to be free-riders, before they have a chance to reap any benefits.

Delton et al. explored this by asking college students to judge fictitious characters who either did or did not partake in a collective task (Delton, Nemirow, Robertson, Cimino, & Cosmides, 2013). The characters could choose to partake in a car wash which raised money which could then be used by the group. In one scenario, the money raised was only used by individuals who actually helped in the carwash (club good), and in the other the money was shared by everyone, regardless who helped (public good). Delton et al found that participants judged those who didn't help in the public good scenario, *even before they partook of any group benefits*. They confirmed this result in an experiment with real monetary rewards, showing that individuals who opted out of public goods were judged most harshly (Delton et al., 2013).

The above results demonstrate, in a lab setting, that evolved psychological mechanisms may exist for regulated cooperative behaviour. But does this have any bearing on cooperative behaviours in contemporary society? Tooby and colleagues suggest it does.

For example, consider the result that foraging effort influences participants' judgement of foragers as free-riders. Tooby and colleagues wanted to know whether this can explain modern views about welfare (Petersen, Sznycer, Cosmides, & Tooby, 2012). People often negatively judge welfare recipients. Is this because they judge them as free riders? If so, then the results from the vignette experiments should correlate with people's real life opinions about welfare.

Petersen et al. used real-life surveys to find out if that was true. They asked people to rate statements from completely disagree to completely agree, on a range of welfare-related topics. From the surveys they determined that it was, specifically, perceptions of welfare recipients effort to find work that drove opinions about welfare,

suggesting that the results from fictitious vignette experiments hold in real, societal circumstances.

What other aspects of society can be explained by evolved cooperation-regulation mechanisms? One aspect of cooperative behaviour is alliance formation – knowing when alliances are forming, detecting who is a part of them, and making decisions about joining them. Modern alliances might include sports fans, political parties, religious groups, and so on. Do humans have designed psychological mechanisms for detecting alliances?

Tooby and colleagues extended their categorisation research to study alliance detection, because if the hypothesis that alliance detection has evolved in humans is true, alliances should form a subset of the categories that humans can use (Pietraszewski, Cosmides, & Tooby, 2014). Earlier research suggested that human subjects can detect when individuals are teaming up in conflict against a third party. But if alliance-detection systems evolved for the purpose of cooperation, alliances should also be detected in the absence of conflict. Further, if these mechanisms are real, alliance detection of a cooperative group should be able to override other, empirically well-established categories, such as race. Pietraszewski et al. tested each of these ideas.

They utilised a version of the categorisation method described above, termed "Who said what", in which subjects are asked to recall, unsurprisingly, who said what. Once again, the mistakes subjects make reveal what categories they have formed. When a subject remembers who said what correctly, there is no way to know what features they used to remember it. But when they make mistakes, if these mistakes are consistently biased in one direction, it can reveal which features they use to form categories. For example, if they mistakenly choose members of the same race as the original "sayer", this suggests some features of race inform their categorisation.

First the researchers used this experimental set up to show that, in the absence of other cues, subjects formed categories based on race, in line with previous results. Next, they added charity group membership to the descriptions of people, with two groups each focused on cooperating to help other people, where these charity groups served as stand-ins for "cooperative alliances".

The researchers found two things. First, subjects formed categories based on the cooperative alliances, grouping together people who donated to the same charities in their minds. Further, in doing so, subjects downregulated their pre-existing racial categories. That is, as subjects began to form categories based on who donated to what charity, the firm racial categorisation lines in their minds began to blur and dissolve. This suggests that we *do* have alliance detecting mechanisms in place for cooperative behaviour. But what, if anything, does it tell us about race?

Well, there's an added twist. Simultaneous to the studies described above, the researchers measured detection based on sex. Their reasoning was as follows: sex is a well-known existing category. However, they argue that it is distinct from race in that it is fundamental to mammalian biology, and has a much older evolutionary history than race. Thus, they predicted that, while race is downregulated, sex categorisation should be unaffected by the detection of the new, charity-group based alliances. This is exactly what happened, suggesting that not only does this study show that we can detect cooperative alliances and readily shift them, but that race, unlike sex, is one of those alliances, and therefore could be shifted as a category (Pietraszewski et al., 2014).

Finally, Tooby and colleagues repeated this study with political parties, instead of charities, to see whether we can understand political parties as a modern example of the kinds of alliances our brains evolved to detect. Using democratic and republican parties as their test case, they recovered their results: human subjects upregulate categorisation based on political party alliances, downregulating categorisation based on race but not sex. This suggests that not only did these mechanisms evolve for the purpose of detecting cooperative alliances, but that political parties may be modern examples of such alliances that humans still use evolved mechanisms to detect (Pietraszewski, Curry, Petersen, Cosmides, & Tooby, 2015).

Trying to be chosen as a partner

The studies above suggest we have evolved psychological systems for choosing partners. But those mechanisms would only evolve if being part of a cooperative group or task was evolutionarily beneficial. And if that's true, then we should have evolved mechanisms for *being chosen*, as well as choosing correctly.

Tooby and colleagues tested this idea in a series of experiments which studied the link between exclusion, emotion, and behaviour (Robertson, Delton, Klein, Cosmides, & Tooby, 2014; Sznycer et al., 2016). Their framework is as follows: exclusion from social groups is costly; exclusion should therefore trigger behavioural responses aimed at either regaining inclusion to the group or minimising the costs of exclusion; these behaviours will be mediated by specific emotions.

In their first study, they presented participants with a series of exclusion scenarios, and then asked them to describe their emotional response and imagined behavioural response. For example, participants would be told that they were part of a basketball team or restaurant staff, and then told, despite a long history of being part of this group, that they had been excluded for one of five reasons: free-riding, inability, disruption to the group's coordinated function, having considered leaving the group, or having an infectious disease. They were then asked to rate their shame, anger, fear, etc. Finally, they were asked to rate their behavioural response, e.g. apologise and ask for forgiveness, work harder, evoke guilt from the group, etc.

The experimenters predicted that specific reasons for exclusion would evoke specific behavioural responses, and that these would be mediated by specific emotions. Overall, that's what they found:

Emotions and strategies predictions and results from Study 1 summary.

		Predictions	Confirmed in Study 1	Unexpected in Study 1
Free riding	Emotions	Guilt, Shame	Guilt, Shame	None
	Strategies	Apologize, Be a team player, Work harder	Apologize, Be a team player, Work harder	None
Inability	Emotions	Anger, Guilt	Anger, some Guilt (less than free riders)	None
	Strategies	Apologize, Remind group of shared history, Work harder	Apologize, Work harder	Low Remind group of shared history
Coordination disruption	Emotions	Guilt, Shame	Guilt, Shame (less than free riders)	Anger, Betrayal
	Strategies	Apologize, Work harder, Be a team player	Apologize, Work harder, Be a team player	None
Exit from the group	Emotions	Guilt, Pride or Satisfaction	Pride (relative to other conditions)	Little Guilt
	Strategies	Apologize, Work harder, Remind group of shared history	Apologize, Work harder, Remind group of shared history	None
Pathogen infection	Emotions	Anger, little Guilt	Anger, little Guilt	None
	Strategies	Remind of shared history, Evoke guilt from group, Work harder	Remind of shared history, Evoke guilt from group, Work harder	None

Figure X. Table 1 from Robertson et al. 2010 PERMISSION NEEDED.

Of course, a weakness of this study is that it simply measured participants *imagined* responses, which might be wholly different from their actual responses. But at the

very least, it suggests a link between exclusion and specific emotion-mediated behaviours, demanding more research.

Tooby and colleagues followed this up with a specific study of shame which was more closely tethered to real scenarios, and which provided a comparative test of the hypothesis, adding power to their analysis. Their hypothesis was that shame is an evolved mechanism to deter individuals from making decisions which will cause them to be devalued, prevent them from telling others about things that would devalue their social standing, and help individuals respond appropriately when they are devalued.

Importantly, this allowed the researchers to make a roughly quantitative prediction. In "deciding" whether to feel shame, one should estimate the magnitude of devaluation, so as not to invoke a costly emotional response unnecessarily. Therefore, shame should scale with the magnitude of devaluation – a stronger prediction then merely "shame" or "no shame".

The researchers found that, across several countries, shame tracks the threat of devaluation from others, lending credence to the hypothesis that shame is evolved mechanism for managing social inclusion (Sznycer et al., 2016).

Punishment

There is no doubt that punishment is an established and complex part of human society, and that it is part of our evolved psychology. But punishment remains one of the more controversial topics in both evolutionary biology and psychology.

Does punishment promote cooperation? One intuition is yes: by reducing the victories of cheaters relative to co-operators, punishment must be good for cooperation. Another intuition says otherwise: punishment itself is costly, and at best would only be favoured under the very same conditions that cooperation itself is favoured, thereby rendering it unnecessary. Does punishment lead to the evolution of cooperation when it wouldn't otherwise evolve? Does it maintain cooperation once cooperation has already evolved? Or is it orthogonal to cooperation, a selfish trait used to increase *personal* gains? Even theoretical models disagree about the role punishment can play, making it all the more difficult for empirical results to be clearly interpreted.

Still, that's not to say there isn't a significant amount of research on the topic. Tooby and colleagues believe they have evidence that punishment, as an evolved psychological mechanism, is *not* for upholding group cooperation (Krasnow, Cosmides, Pedersen, & Tooby, 2012). Instead, they argue, it's used for personal gains.

Krasnow et al. constructed in experiment in which participants, playing online games, had opportunities to trust, cooperate with, and punish partners. The participants had information about either the partner's reputation for cooperating with a third party, or cooperating with a first party, or both. Participants could choose to gain real sums of money by cooperating with their partner, which required trusting that the partner would share the financial gains.

Krasnow et al. found that while participants used third party reputation information in deciding whether to trust their party when it was the only information available to them, they did not utilise it when there was also first party information. This is consistent with participants using reputation information to make decisions about how they will be affected personally, but not consistent with them using reputation information about how others will be affected.

Further, they found that individuals were more likely to cooperate with a partner they had previously punished, suggesting that a role of punishment is to increase personal gains from interactions (Krasnow et al., 2012).

Tooby and colleagues elaborated on this result with an extensive third-party punishment game experiment. In this experiment, participants played three player games, in which one individual acts as "dictator", another as a "recipient", and a third as a "punisher". The dictator is given two tasks. First, they are given a sum of money, which they can choose how to divide up between themselves and a recipient. They can keep it all and give nothing to the recipient, or vice versa, or somewhere in between. Second, they are given an opportunity to receive a gift of money or deny it, where in the case of denying it, either the recipient or the punisher receives a gift of money. For example, they can either accept five dollars, or reject it, in which case a recipient will receive twenty dollars.

Meanwhile, punishers are given a sum of money, which they can use to "punish" the dictator. They can spend none, some, or all of their money to take money away from the dictator. For example, if the dictator doesn't give any money to the recipient, the punisher can pay increasing amounts to have money taken away from the dictator, even though that money won't return to anyone in the game.

The experimenters then asked the punishers to make a number of judgements (before the game was played). First, given a range of possible divisions the dictator could make (e.g. $0/$10, $4/$6, $10/$0), what subsequent trade-off would they be willing to make. That is, if a dictator is the kind of person who would keep all 10 dollars, would they then be willing to forego a dollar in order for the punisher to be given a hundred dollars? The punisher must consider all possible behaviours by the dictator and make judgements about each. Further, they were asked, given a dictator's behaviour towards a recipient in the division task, what their behaviour would be towards the punisher in the trade-off task.

Their prediction was as follows: if punishment is *for* personal gain then punishers should a) predict a correlation between how the dictator divides up the money given to her in the first task (the division task) and how much she is willing to forego to benefit the recipient in the second (trade-off) task, b) predict a correlation between the dictator's behaviour towards recipient and themselves (i.e. that if the dictator treats the recipient badly, she will treat the punisher badly, too), c) make decisions about punishing accordingly to (a) and (b), and, importantly, d) punish less when they have been treated well, regardless of how a third party has been treated. They found that all of these predictions were upheld, and claimed the most parsimonious explanation for all of these results is that punishment is for personal gain, not upholding social norms (Krasnow, Delton, Cosmides, & Tooby, 2016).

One criticism of this approach is that it uses something called the "strategy method", in which the participants are asked the make decisions about their actions and feelings prior to actually experiencing anything. However, other experiments have looking at the impact of the strategy method, versus real responses, and found it not to affect these kinds of third party punishment games (J. Jordan, McAuliffe, & Rand, 2016).

A broader concern for these studies is whether participants truly understand the games being played, and therefore whether their actions reflect true preferences (Burton-Chellew et al., 2016). Krasnow et al. did have participants complete comprehension questionnaires prior to playing. However, a preferable metric would be whether their actual game play, under circumstances where the "correct" behaviour is unambiguous, is consistent.

Turning to the world outside the lab, we might ask whether these studies on how individuals punish in the lab have any bearing on how humans behave naturally. An obvious arena to study punishment is in the legal system. In our justice system there are two fundamentally different responses we can take when a crime is committed: punish or repair.

Tooby and colleagues argue that how individuals respond to crimes is driven by how much value they attribute to criminals. Specifically they argue that there are two different axes on which a person responds when confronted with a crime: a seriousness axis, which determines how *much* punishment an individual should receive, and a social value axis, which determines *whether* an individual should be punished (or given an opportunity to rehabilitate). They argue that this latter exists because if an individual is attributed high enough social value, then it is personally beneficial if that individual remains a functioning part of the group.

They studied this using surveys of secondary and university level students in Denmark and the USA. They were shown vignettes of different crimes, such as robbery and rape, committed by members of society with different potential "social value" (e.g. "he owns an autoparts store in town"). They were asked questions aimed to determine what social value the participant attributed to the person (e.g. "how likely is it this person will contribute to society in the future"), how serious they thought the crime was, whether the person should be punished or rehabilitated, and what punishment was appropriate.

They found, as predicted, that the perceived seriousness of the crime correlated with the magnitude of the punishment, and that the perceived social value of the person correlated with whether the punished should be punished or repaired (Petersen, Sell, Tooby, & Cosmides, 2012).

However, these results have a broader significance as well. It's very unlikely, in the modern world, that an individual would be affected by whether a specific, random criminal is punished or rehabilitated. However, that would not have been true thousands of years ago, when we lived in small, isolated groups. Thus, this result, along with the many others described here, further suggest that we have fine-tuned, precise psychological systems, designed by selection during our long evolutionary history, for the purpose of managing cooperative tasks and behaviour.

The origins of cooperation in humans

Evolutionary psychology has unearthed increasing evidence that the very complex cooperative systems we have in place today – police, schools, governments – are powered, at least in part, by cooperative adaptations in the brain. But how do such systems emerge? It's not a small step from communally eating a carcass to drawing up a constitution. Is there evidence for complex cooperative systems?

Lee Cronk has spent the last decade studying exactly that: a relatively simple, but still uniquely human, cooperative system in the real world. The cooperative "game" he studies is called Osotua, a reciprocal gift given practice engaged in by the Maasai of East Africa.

An Osotua gift is a gift of wealth, usually in the form of cattle (because that is the Masaai's main currency). Once given, it an establishes an Osotua bond which lasts for life – in fact, it is passed to children if the Osotua is not fulfilled before death. The bond formed is that the gift receiver must return the favour to the giver.

However, there are several unique features of this relationship. First, the origin gift comes upon request, and the request is only made based on *need*. Second, the obligation formed is to help the original giver when they are also in *need*. Thus it is a need based transfer. Third, the return gift need not be of the same magnitude or quality as the original gift. Thus it is fundamentally distinct from debt, which Maasai also have, and which may be utilised for things other than need and requires reciprocal repayment. This makes Osotua a unique and unusual form of wealth exchange, and also a potentially valuable test case form primitive cooperative systems.

Lee Cronk set out to discover the purpose of Osotua, or "need based transfers" more generally. Is the gift altruistic or a form of reciprocity? Is it favoured because of the benefits accrued to relatives or self?

In 2007 Cronk demonstrated experimentally that Osotua was something fundamentally different from standard reciprocity games (Cronk, 2007). He had Maasai individuals play a simple trust game. In a trust game, two players are given some money. The first player can then give all, some, or none of his money to the second player, which causes that amount of money to be tripled. The second player can then return any amount of money to the first player. However, the twist was that with half the players they were simply told at the start "this is an Osotua game".

In the standard treatment, individuals followed normal reciprocity rules. The amount they gave to their partner was proportional to how much they expected to get back, etc. However, in the "Osotua" treatment, the simple framing of the game was enough to change the way people played. In this case, how much they gave to the partner was *negatively correlated* to how much they expected to get back.

This led Cronk to hypothesise that Osotua was a risk-pooling system. Risk-pooling is a special kind of cooperation, in which individuals transfer their risk to the group, to minimise the *average risk* experienced by a random individual in the group.

Cronk and colleagues first tried to determine, at least in principle, whether an Osotua like system could work to minimise risk in this way and benefit the group. To do this, they developed computer simulations of individuals raising herds of cattled. They then directly compared the success of a) an Osotua system b) a system of no cattle exchange and c) a system of random, probabilistic exchange of cattle. They found that Osotua significantly increases herd longevity via pooling and minimising risk (C. A. Aktipis, Cronk, & de Aguiar, 2011).

One way in which Osotua is a valuable discover is that it provides a real life example of a relatively simple cooperative game. Why is this useful? A major criticism of all of the cooperative game experiments described so far is that they are highly artificial. Of course they are meant to get at more fundamental principles of cooperation in everyday life, but there is always a possibility that the games are so abstracted that they don't measure how people really behave. The problem is that real life cooperative behaviours (like a legal system) are too complicated to import into the lab.

Cronk believes Osotua gets around this to some degree by providing a real life behaviour that is simple enough to mimic in the lab ((Gazzillo, Sopher, Aktipis, & Cronk, 2013), unpublished). Cronk and colleagues had individuals play an Osotua like game in the lab, in which they were given a resource that would grow and shrink over time (like a herd), where of course the "time" in this case was condensed into a series of rounds. Occasionally "shocks" (similar to droughts or disease) could drastically reduce the resource. They also had opportunities to exchange wealth with their partners.

Some subjects were read briefs about Osotua, others Western American cattle farming, and others nothing. They then used advanced statistical methods to tease out the players' behaviour. Do they give more when there is greater inequity between them and their partner? When their partner has given them more in the past? And so on.

Gazzillo et al. claim to show that individuals do not show "inequity" aversion, and instead utilise sharing for selfish reciprocity. Further, while subjects tolerated some degree of cheating ("over-harvesting" the resource), too much cheating would lead to someone being cut-off. While these results themselves may not be particularly surprising, the key advance is in identifying a potentially more "ecologically" realistic way of studying human behaviour in lab games.

Next, to further understand how Osotua might work to reduce risk in principle, Cronk and colleagues extended their 2011 model to include explicit networks of people and more precise rules governing who asked who for an Osotua gift (Hao, Armbruster, Cronk, & Aktipis, 2015). They varied: the size of the network, the connectivity of the network (how many other people an individual could interact with), the heterogeneity of the network (essential how uniform the connections are), and whether individuals selectively asked the wealthiest person for Osotua or asked at random. Their goal was to find out which form of Osotua was optimal, in order to then ultimately compare to what they found in the field.

The results were somewhat complicated. Increased connectivity always led to greater herd survival. Larger network size led to greater herd survival, but only when individuals selectively asked for help from the wealthiest individuals. Increased heterogeneity (e.g. one individual is connected to two others while another is connected to ten others) *increased* longevity when partners were selective but had the opposite effect when they asked for help at random. Generally selective asking increased longevity, so it would be interesting to see if that is how Osotua is done in practice (Hao et al., 2015).

In a second extension to the model, Aktipis et al. directly compared Osotua to a system of debt. With Osotua, requests for help are only made according to *need*, gifts are given as long as the requested can do so without harming their own survival, and Osotua return gifts need not match the original gift (no account keeping). This is opposed to debt, in which accounts are kept, exchanges must match each other in value, and gifts might be requested for purposes other than absolute need.

Interestingly, they found that an Osotua-like system of need-based transfers outperformed a debt-like accounting system. Osotua led to greater pooling of individual risks, higher herd survival, and even greater wealth equality (A. Aktipis et al., 2016). However, one question for this type of modelling going forward is: how relevant a measure is herd longevity? Ultimately what matters is its impact on fitness, and therefore an important test would be whether herd longevity is the appropriate proxy.

Regardless, Osotua is clearly an effective risk-pooling mechanism, and a shining example of an early cooperative system, potentially similar to early systems during the origin of advanced human cooperation. An obvious question, however, about Osotua and similar risk-pooling mechanisms, is: what prevents cheating? For example, what stops someone from receiving a gift and then never returning it?

Cronk and Aktipis have argued that the threat of supernatural punishment may provide one answer. The Osotua bond is "sacred", suggesting that violating it will lead to punishment more severe than the benefits gained from cheating (Cronk &

Aktipis, 2018). Whether this is the case requires further empirical work, and if it turns out to be true, raises the question of what maintains belief in the supernatural punishment, and why that emerged in the first place.

The How: The mechanism of cooperation

In the beginning of this review, we talked about Tinbergen's Four Questions (Figure 1 repeated here for convenience), and about how we would be focusing on the "why" questions.

	Dynamic	**Static**
Proximate	Ontogeny (Development)	Mechanism (Causation)
Ultimate	Phylogeny (Evolution)	Function (Adaptation)

In fact, we really mostly focused on the ultimate why question: why, or for what purpose, did cooperation evolve? But now we turn, briefly, to the how question. How do we cooperate? Specifically, what *mechanisms* manage cooperation in humans. How does are brain decide when to cooperate and when to be selfish? What chemical systems regulate this? How, if at all, is cooperation regulated differently from other mental behaviours? We'll start from the outside of the brain, looking at *inferred mental processes*, and turn to the meaty inside afterwards.

A cooperative mind

How are cooperative decisions made? Are we inherently selfish, but upon slow deliberation of possible outcomes, choose to be occasionally cooperative? Or is it the other way around: cooperation is our knee-jerk reaction, only occasionally fettered by carefully thought through selfish decisions?

Daniel Kahneman and Amos Tversky were among the first to point out that that we seem to have a dual-processing brain that works on two speeds. A fast brain, which makes quick decisions without "thinking" or deliberation (what we might

call our "intuition") and a slow brain that moderates the fast one with deliberation and logic. Where do cooperation and selfishness fit in to this schema, if at all?

David Rand, Martin Nowak, and colleagues have set out to untangle this question with a series of experiments measuring cooperative and selfish behaviour, along with reaction times and other factors that might influence "fast" versus "slow" thinking.

In 2012, Rand et al. measured deliberation times – or how long someone took before behaving – and behaviour across a series of experiments (David G Rand, Greene, & Nowak, 2012). They found several results: a) individuals who decide more quickly tended to be more cooperative, b) when experimenters force subjects to decide quickly it increases cooperation, and c) when experimenters tell subjects to "trust their intuitions" it increases cooperation. Rand et al. concluded that, taken together, these findings suggest that cooperation is our intuitive, spontaneous response, whereas selfishness is the product of slower deliberation. Further, this work was supported to some degree by theoretical models, which showed, in a simple game theory model, that while intuitively cooperative individuals who modify their behaviour with selfish deliberation can evolve, the inverse does not (Bear & Rand, 2016).

However, this work was criticised for excluding individuals who took too long to respond (which might bias the results) and controlling for time in one of their statistical analyses, potentially obscuring the interaction between time and cooperation (Tinghög et al., 2013). While Rand et al. then redid the analysis, claiming to still find the same, original results, other similar studies actually found the opposite result: increased deliberation was associated with *more* cooperation, not less. What was going on?

Rand, Nowak, and colleagues developed a theory that might explain these conflicting results: the "social heuristics hypothesis" (David G. Rand, Peysakhovich, et al., 2014). Their idea is as follows: our fast, intuitive responses are those that have typically been "successful" in the past. In other words, if certain behaviours have consistently led to rewards (material, existential, etc.), those will be the behaviours we internalise as our default, intuitive reactions. However, those behaviours won't *always* be successful, in all situations. Thus, when given the opportunity, we can use careful, slow deliberation to moderate our behaviour to match the current situation. Further, they argue that because cooperation is usually successful in our everyday lives, that will usually be the quick, intuitive response.

They tested this hypothesis by looking across 15 time pressure, cooperation game experiments, as well as devising an additional, novel time pressure experiment. On average, across all studies, they *did* find a positive effect of time pressure (decreased time for making decisions) on cooperation, suggesting that cooperation is our default move. However, within studies there was substantial variation, and the effect wasn't so clear cut. However, they found that this variation can, to some degree, be explained by previous experience in lab games. That is, those who had played previous games in which cooperation *was not* successful had "learned" to have quicker, intuitive *selfish* responses, which is exactly what the social heuristics hypothesis would predict (David G. Rand, Peysakhovich, et al., 2014).

Further, Rand and colleagues have suggested that simple interpretations of reaction times may be misguided (Evans, Dillon, & Rand, 2015). They argue that reaction times are not a good proxy for whether someone is using intuitive or reflective thinking. Instead, "conflict" may drive reaction times, with extreme decisions, where the correct response is obvious, occurring quickly, and not

necessarily aligning with intuitive decisions. Thus, they argue, reaction times on their own are misleading, and should, at the very least, be combined with primes for intuition or deliberation (Evans et al., 2015).

Guided by this result, Rand conducted a meta-analysis of 67 studies which measured the effect of "cognitive-processing manipulations" (e.g. starting the experiment by saying "try to trust your intuitions here") on cooperation. Rand argued that the social heuristic hypothesis predicts different effects of such manipulations on "pure cooperation" versus "strategic cooperation". Pure cooperation, he argued, is when one cooperates in a one-shot interaction in which there is no scope for benefits returning to the individual, whereas strategic cooperation can include cooperating in reciprocity games to increase one's own ultimate winnings. He argues that deliberation will decrease pure cooperation, because it allows slow, selfish thinking to prevail, while it will have no effect on strategic cooperation, which already invokes deliberative thinking (David G. Rand, 2016). Indeed, this is what Rand found: an increase in pure cooperation when intuition was promoted by the experimenters, but no effect of intuition/deliberation on strategic cooperation.

However, the above results focus on a mixture of one shot and repeated interactions. Most real-life interactions aren't one shot, though. What happens when we look exclusively at repeated interactions?

Rand and colleagues argue that social heuristics predicts that, in this case, the social environment will determine whether cooperation or selfishness occurs more quickly. Specifically, they argue that we internalise the norm behaviours of our social environment – in this case, what our partners are doing. And thus, it will be reciprocal behaviour – matching our partner – that is fast, not cooperation or selfishness per se. They looked across four different studies, including games in which cooperation was more successful and in which selfishness was more successful. They found that reciprocal behaviours were consistently fastest, and thus, it was the social environment that determined the speed of behaviour. Further, they showed that this was driven by conflict, and thus does not reveal which behaviour is "intuitive" or "deliberative" (Nishi, Christakis, Evans, O'Malley, & Rand, 2016). However, in moving forward, and potentially designing an experiment that measures real intuition, this study shows that reciprocity may be the key factor to look at.

Finally, Rand and colleagues were interested in the role of gender on shaping whether cooperation was intuitive. The social heuristics hypothesis predicts that previous experience will shape current intuition. Research suggests that women are more often expected to be cooperative than men, which means that indirect benefits of cooperation via indirect reciprocity may more often reinforce cooperation as a successful behaviour for women (David G Rand, Brescoll, Everett, Capraro, & Barcelo, 2016). Further, women are more often in roles (e.g. being a mother) that require cooperative behaviour (David G Rand et al., 2016). Rand et al. suggest that this means cooperation may be more intuitive for women than men.

This prediction was supported by their meta-analysis: priming subjects to focus on intuition over deliberation led to increased cooperation for women but not men. Further, the more female subjects identified with traditionally masculine terms, the more deliberation increased their selfishness (David G Rand et al., 2016). These results support the social heuristics hypothesis *and* a role for gendering in shaping the intuitiveness of cooperation.

A potentially open question for researchers interested in the speed of cooperative decision-making is whether cooperative or selfish moves in the games

being studied inherently require more mental steps. In some cases, for example, being strictly selfish requires a number of calculations, whereas there are many simple ways to "cooperate". It would be interesting to see what effect controlling for this variable has on the existing results.

The Chemistry of Cooperation

One way to learn about the mechanistic side of cooperation is to observe behaviour and make inferences about what's happening in the mind, as we saw above. But another approach is to actually physically manipulate the brain, to learn about what physical, biological factors influence cooperation.

A number of researchers have done exactly that. Oxytocin is a chemical produced by the body that is known to play a role in promoting pro-social behaviours, such as cooperation, in humans and other primates. In 2015, however, Brosnan et al. found a surprising result: increased Oxytocin actually had a negative effect on food sharing.

The researchers designed an experiment to study Capuchin monkeys, a species known to be highly social, and which have been shown to handle social dilemmas in a manner not entirely distinct from humans (Brosnan et al., 2017). The capuchins were either injected (through the nose) with Oxytocin or an inert placebo. The capuchins were captive, socially living monkeys, and were observed in pair groups. After receiving the nasal spray, the one of the monkeys could gain a food reward, and the possessor of the food could either actively share it with the partner or the partner could take some of the food itself (passive sharing) (Leverett et al., 2015).

Interestingly, Oxytocin had no effect on whether the monkeys actively shared their food. But when either the possessor of the food or the recipient received Oxytocin, it actually decreased passive sharing. The researchers also found that it decreased the amount of time the monkeys spent in proximity to each other. This reduction in sharing and proximity came as a surprise, because previously Oxytocin was known to increase cooperation.

How can we explain this puzzling result? One possibility is that the reduction in proximity, and therefore passive sharing, was a red herring of sorts. Perhaps, the researchers surmised, Oxytocin is linked to a *short term* reduction in proximity, even while it increases sociality over the long term.

To study this, Benitez et al. devised a new approach to studying Oxytocin: measuring the amount of Oxytocin in urine (Benítez, Sosnowski, Tomeo, & Brosnan, 2018). To do this, they first confirmed that the amount of Oxytocin present in capuchin urine was actually a good proxy for how much Oxytocin was in their system, which they did by spraying Oxytocin into their noses and then measuring their urine. The results showed that, indeed, urinary Oxytocin levels were a good measure of how much Oxytocin the monkeys had in their bodies.

Next, they measured urinary Oxytocin levels during and after grooming and fur rubbing events. Grooming and fur rubbing are known to be mechanisms by which capuchins generate and strengthen social bonds. The researchers found that Oxytocin levels increased during such grooming events. Interestingly, however, after these grooming events the monkeys immediately decreased their proximity to each other, for 15-30 minutes after the grooming. But, over long periods of time, the grooming events ultimately correlated with increased proximity and other "affiliative" behaviours.

The conclusion, then, is that Oxytocin does play a role in positive, pro-social behaviours. However, for some reason, it is linked with a short term reduction in sociality, only to precipate a longer term, positive effect (Benítez et al., 2018). The exact role of Oxytocin in this context remains unclear, but together these studies confirm that it *does* play *some kind* of important role in regulating social behaviour. One possibility the researchers propose is that the immediate reduction of sociality after Oxytocin production is related to Oxytocin's ability to reduce stress, which lowers the individuals' need to be in proximity to each other. Perhaps Oxytocin production is the body's way of telling the capuchin that it has reached its quota of socialising for now, and can take a break – but for now that remains mere speculation.

Oxytocin doesn't just impact cooperation in capuchins – abundant research has demonstrated a similar role for Oxytocin in humans. However, much of this research has been on how groups treated with Oyxtocin differ from groups treated with a placebo. A more powerful approach is to actually look at how an individual's behaviour changes after being treated with Oxytocin.

That's exactly what Chen et al. (2017) did (Chen, Gautam, Haroon, & Rilling, 2017). Using a double-blind approach, they had individuals play a prisoner's dilemma game. Some individual's were given a nasal placebo, while others received a spray of Oxytocin. Interestingly, while the effects of Oxytocin identified in between-group studies were recovered in women, these effects no longer seemed present in men. Hecht et al. (2017) extended this work by looking more carefully exclusively at women (Hecht, Robins, Gautam, & King, 2017). This study suggested that Oyxtocin was linked to reduced social attention in women.

Finally, there is the question of how this actually plays out in the brain. Is there a cooperation centre of the brain? Where in the brain does Oxytocin act? A team of researchers determined this by dissecting, dyeing, and imaging brains of humans, rhesus macaques, and chimpanzees. There question was whether Oxytocin works via the cortex, a part of the brain that plays a critical role in cognition. To find out, they essentially washed the cortex with Oxytocin antibodies: the prediction being that if fibers in the cortex were designed to read and respond to Oxytocin, these antibodies would bind to them. After washing the brains, imaging revealed that while the rhesus cortexes had no Oxytocin-related fibers, both humans and chimpanzees had cortical fibres that were Oxytocin-immunoreactive, suggesting that Oxytocin effects act at least partially through the cortex (Rogers et al., 2018).

Conclusion

The puzzle of cooperation revisited

Is cooperation still a puzzle? In some fundamental ways, the answer is no. Darwin's migraine over the sacrificial bee can be assuaged with the help of genetical relatedness. An organism will adopt a cooperative behaviour if the benefits of that cooperation are targeted preferentially at gene-sharing relatives. The bee, in stinging an intruder, gives up one whole set of its genes, in exchange for hundreds of half-sets of its genes – a more-than-fair trade in the eyes of selection.

This idea was formalised by Hamilton in 1964, and understood most straightforwardly through the simple dictum (Hamilton's rule): a trait will spread if the benefits of the trait (b), weighted by relatedness to beneficiaries (r), outweigh the costs to the trait's bearer (c), i.e. $rb – c > 0$. Hamilton's rule, and the more general concept of inclusive fitness (the measure of fitness that takes into account social effects) has spawned a great research programme, guiding empirical work across the tree of life, from microbes to mammals. While evolutionary biologists believe genetic relatedness accounts for some cooperation in humans, adaptive hypotheses are hard to test in humans, because we cannot easily experimentally manipulate our biology or ecology.

Certain exemplar study systems, such as *Dictyostelium discoideum*, have allowed researchers to empirically test the role of relatedness in cooperation, study the evolutionary dynamics of cooperation and cheating, and even detect signatures of social selection in the genome. On the other end of the scientific spectrum, theoreticians have developed new ways of incorporating the effect of relatedness in mathematical models, for example by capturing features of population structure and relatedness in simple graphs.

Genetic relatedness and Hamilton's inclusive fitness theory resolve Darwin's puzzle of the bee's ultimate sacrifice. But, more generally, cooperation can evolve whenever the cooperative trait correlates, on average, with higher fitness. In the case of genetic relatedness, this happens because while some individual bearers of the trait suffer the costs of cooperation, on average across the population, co-operators do better, because the co-operators interact with each other more than by chance (in other words, *relatedness* is a measure of assortment). This is cooperation due to *indirect benefits*. But the benefits of cooperation might also accrue *directly to the co-operator*, for example because of reciprocity.

This latter form of cooperation is the one that has been widely studied in humans, for two reasons: first, the costs and benefits can be measured within a single generation, or even a single interaction, and second, interesting forms of direct benefits arise with memory, learning, emotions, culture and institutions.

Growing empirical evidence suggests that humans have evolved psychological mechanisms to regulate cooperation, that cooperative behaviours vary across individuals and cultures, and that institutions, such as religion, can impact individual cooperation. Further, researchers have begun to identify the mechanistic causes of cooperation in the primate brain, revealing Oxytocin as a crucial chemical

in regulating social behaviours. As a whole, this body of work is painting a picture of both the ultimate (adaptive) and the proximate (mechanistic) causes of cooperative behaviour in humans.

Open questions in cooperation research

But does this mean the puzzle of cooperation is completely solved? Not exactly. While inclusive fitness theory provides a general explanation for adaptive cooperation, for decades critics have pointed out that models of inclusive fitness require certain assumptions, such as additivity of fitness effects. Proponents have argued that either this isn't strictly true, or that the deviations are random and unbiased, and of small magnitude, and therefore don't significantly affect the predictions about when cooperation will evolve and to what degree.

Critics of inclusive fitness theory have yet to provide a viable alternative, especially one that can be used by biologists doing empirical work. Will future theoreticians, using some new mathematical technique, identify a "truer" maximand than inclusive fitness, which makes fewer assumptions? Or will social biologists find ways to formalise the subtle verbal description of inclusive fitness, allowing biologists to proceed in using inclusive fitness, assured that the theory rests on firm theoretical grounds? A third possibility is that there is no superior maximand, but inclusive fitness will never be able to predict behaviour across all settings. Perhaps natural selection only pushes organisms towards the target of inclusive fitness, missing in unpredictable and therefore unhelpful ways, and biologists will be forced to accept an imperfect predictor.

Microbes are pushing these frontiers of cooperation research, for two reasons. First, they violate many of the norms of the higher organisms for which evolutionary theory was largely built in the 20th century. They have high mutation rates, exchange genetic information horizontally as well as vertically, and often lack mechanisms to reduce internal genetic conflicts. Second, they lend themselves to empirical testing. They are easy to keep in the lab and can undergo evolution over short human-timescales. Many of the successful empirical tests of inclusive fitness theory have been carried out in microbes in the last two decades. What will they tell us next about the murky fringes of cooperation theory, such as the roles of non-additive fitness effects, multiplayer games, and strong selection?

And finally, there is the last great frontier of cooperation: humans. Humans are the hardest to study, because many adaptive experiments are unethical in humans, and opportunities for comparative tests are few and far between. But we also possess traits which make our cooperation particularly interesting (albeit not particularly extreme, compared to ants and wasps): memory, consciousness, emotions. Perhaps most importantly, we are strongly shaped by cultural evolution, and yet the fundamental theory of cultural selection is under-formed. What kinds of behaviours can cultural selection mould? What is the relationship between genetic and cultural evolution?

The use of new tools for studying human cooperation, like economic games, has spread rapidly. However, the controversy over how to interpret these games, and when they are convincing evidence for prosocial behaviour, belies how complex and challenging humans are as a subject. Still, advances in game theory models, experimental design, and technology for studying humans over their lifetimes and across populations, suggest a bright future for the field.

It is a pet peeve among some social biologists that so many articles about cooperation start with some version of the phrase "cooperation is a puzzle". To them, cooperation *was* a puzzle, but theories of inclusive fitness and reciprocity resolved it. And yet, it is these very same social biologists who are *still* dedicating their lives to the study of cooperation. That's because, although people like Bill Hamilton resolved the fundamentals of cooperation, it's the meaty details – why this bird gave up its meal on this specific instance, or why this tribe was more prosocial than the other – that continue to get scientists out of bed in the morning.

References:

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A. C., Andersson, M., … others. (2011). Inclusive fitness theory and eusociality. *Nature*, *471*(7339), E1–E4.

Adlam, B., & Nowak, M. A. (2014). Universality of fixation probabilities in randomly structured populations. *Scientific Reports*. https://doi.org/10.1038/srep06692

Adu-Oppong, B., Queller, D. C., & Strassmann, J. E. (2014). Is there specificity in a defensive mutualism against soil versus laboratory nematodes, Dictyostelium discoideum farmers and their bacteria? *Evolutionary Ecology Research*.

Aktipis, A. (2016). Principles of cooperation across systems: From human sharing to multicellularity and cancer. *Evolutionary Applications*. https://doi.org/10.1111/eva.12303

Aktipis, A., De Aguiar, R., Flaherty, A., Iyer, P., Sonkoi, D., & Cronk, L. (2016). Cooperation in an Uncertain World: For the Maasai of East Africa, Need-Based Transfers Outperform Account-Keeping in Volatile Environments. *Human Ecology*. https://doi.org/10.1007/s10745-016-9823-z

Aktipis, C. A., Cronk, L., & de Aguiar, R. (2011). Risk-Pooling and Herd Survival: An Agent-Based Model of a Maasai Gift-Giving System. *Human Ecology*, *39*(2), 131–140. https://doi.org/10.1007/s10745-010-9364-9

Allen, B., Gore, J., & Nowak, M. A. (2013). Spatial dilemmas of diffusible public goods. *ELife*. https://doi.org/10.7554/eLife.01169

Allen, B., & Nowak, M. A. (2013). O Brave New World with Such Games. *Science*. https://doi.org/10.1126/science.1241750

Allen, B., & Nowak, M. A. (2014). Games on graphs. *EMS Surv. Math. Sci.*, *1*(1), 113–151.

Allen, B., Nowak, M. A., & Dieckmann, U. (2013). Adaptive Dynamics with Interaction Structure. *The American Naturalist*. https://doi.org/10.1086/670192

Allen, B., Nowak, M. A., & Wilson, E. O. (2013). Limitations of inclusive fitness. *Proceedings of the National Academy of Sciences*, *110*(50), 20135–20139.

Athena Aktipis, C., Boddy, A. M., Jansen, G., Hibner, U., Hochberg, M. E., Maley, C. C., & Wilkinson, G. S. (2015). Cancer across the tree of life: Cooperation and cheating in multicellularity. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2014.0219

Bear, A., & Rand, D. G. (2016). Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1517780113

Benítez, M. E., Sosnowski, M. J., Tomeo, O. B., & Brosnan, S. F. (2018). Urinary oxytocin in capuchin monkeys: Validation and the influence of social behavior. *American Journal of Primatology*. https://doi.org/10.1002/ajp.22877

Berbesque, J. C., Wood, B. M., Crittenden, A. N., Mabulla, A., & Marlowe, F. W. (2016). Eat first, share later: Hadza hunter-gatherer men consume more while foraging than in central places. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2016.01.003

Birch, J. (2014). Hamilton's rule and its discontents. *British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axt016

Birch, J., & Okasha, S. (2015). Kin selection and its critics. *BioScience*. https://doi.org/10.1093/biosci/biu196

Brock, D. A., Callison, W. É., Strassmann, J. E., & Queller, D. C. (2016). Sentinel cells, symbiotic bacteria and toxin resistance in the social amoeba Dictyostelium

discoideum. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2015.2727

Brock, D. A., Douglas, T. E., Queller, D. C., & Strassmann, J. E. (2011). Primitive agriculture in a social amoeba. *Nature*. https://doi.org/10.1038/nature09668

Brock, D. A., Jones, K., Queller, D. C., & Strassmann, J. E. (2016). Which phenotypic traits of Dictyostelium discoideum farmers are conferred by their bacterial symbionts? *Symbiosis*. https://doi.org/10.1007/s13199-015-0352-0

Brosnan, S. F., Price, S. A., Leverett, K., Prétôt, L., Beran, M., & Wilson, B. J. (2017). Human and monkey responses in a symmetric game of conflict with asymmetric equilibria. *Journal of Economic Behavior and Organization*. https://doi.org/10.1016/j.jebo.2017.07.037

Burton-Chellew, M. N., El Mouden, C., & West, S. A. (2016). Conditional cooperation and confusion in public-goods experiments. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1509740113

Buss, L. W. (1987). *The evolution of individuality* (Vol. 201). Princeton University Press Princeton.

Buttery, N. J., Smith, J., Queller, D. C., & Strassmann, J. E. (2013). Measuring Cheating, Fitness, and Segregation in Dictyostelium discoideum (pp. 231–248). Springer.

Cavalli-Sforza, L. L., & Feldman, M. W. (1978). Darwinian selection and ``altruism''. *Theoretical Population Biology*, *14*(2), 268–280.

Chen, X., Gautam, P., Haroon, E., & Rilling, J. K. (2017). Within vs. between-subject effects of intranasal oxytocin on the neural response to cooperative and non-cooperative social interactions. *Psychoneuroendocrinology*. https://doi.org/10.1016/j.psyneuen.2017.01.006

Cosmides, L., & Tooby, J. (2013). *Evolutionary Psychology: New Perspectives on Cognition and Motivation*. SSRN. https://doi.org/10.1146/annurev.psych.121208.131628

Cronk, L. (2007). The influence of cultural framing on play in the trust game: a Maasai example. *Evolution and Human Behavior*, *28*(5), 352–358. https://doi.org/10.1016/j.evolhumbehav.2007.05.006

Cronk, L. (2017). Culture's influence on behavior: Steps toward a theory. *Evolutionary Behavioral Sciences*. https://doi.org/10.1037/ebs0000069

Cronk, L., & Aktipis, A. (2018). Sacredness as an implied threat of supernatural punishment: the case of need-based transfers. *Religion, Brain and Behavior*. https://doi.org/10.1080/2153599X.2017.1302978

Darwin, C. (1859). *The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life*. John Murray.

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An introduction to behavioural ecology 4th edition*. Oxford: Wiley-Blackwell.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

Delton, A. W., Cosmides, L., Guemo, M., Robertson, T. E., & Tooby, J. (2012). The psychosemantics of free riding: Dissecting the architecture of a moral concept. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/a0027026

Delton, A. W., Nemirow, J., Robertson, T. E., Cimino, A., & Cosmides, L. (2013). Merely opting out of a public good is moralized: An error management approach to cooperation. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/a0033495

Delton, A. W., & Robertson, T. E. (2012). The Social Cognition of Social Foraging: Partner Selection by Underlying Valuation.

https://doi.org/10.1016/j.evolhumbehav.2012.05.007

Delton, A. W., & Sell, A. (2014). The Co-Evolution of Concepts and Motivation. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721414521631

Dennett, D. (1995). *Darwin's Dangerous Idea*. New York: Simon and Schuster.

DiSalvo, S., Brock, D. A., Smith, J., Queller, D. C., & Strassmann, J. E. (2014). In the social amoeba Dictyostelium discoideum, density, not farming status, determines predatory success on unpalatable Escherichia coli. *BMC Microbiology*. https://doi.org/10.1186/s12866-014-0328-x

Disalvo, S., Haselkorn, T. S., Bashir, U., Jimenez, D., Brock, D. A., Queller, D. C., … Mccutcheon, J. P. (n.d.). Burkholderia bacteria infectiously induce the proto-farming symbiosis of Dictyostelium amoebae and food bacteria. https://doi.org/10.5061/dryad.pb2sq

Ellingsen, T., Herrmann, B., Nowak, M., Rand, D. G., & Tarnita, C. (2012). *Civic Capital in Two Cultures: The Nature of Cooperation in Romania and USA*. *SSRN*. https://doi.org/10.2139/ssrn.2179575

Epstein, Z., Peysakhovich, A., & Rand, D. G. (2016). The Good, the Bad, and the Unflinchingly Selfish. In *Proceedings of the 2016 ACM Conference on Economics and Computation - EC '16*. https://doi.org/10.1145/2940716.2940761

Evans, A. M., Dillon, K. D., & Rand, D. G. (2015). Fast But Not Intuitive, Slow But Not Reflective: Decision Conflict Drives Reaction Times in Social Dilemmas. https://doi.org/10.1037/xge0000107.supp

Everett, J. A. C., Haque, O. S., & Rand, D. G. (2016). How Good Is the Samaritan, and Why?: An Experimental Investigation of the Extent and Nature of Religious Prosociality Using Economic Games. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550616632577

Foster, K. R. (2009). A defense of sociobiology. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 74, pp. 403–418).

Frank, S. A. (1998). *Foundations of social evolution*. Princeton University Press.

Gardner, A. (2017). The purpose of adaptation. *Interface Focus*, *7*(5), 20170005.

Gardner, A., & Grafen, A. (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology*, *22*(4), 659–671.

Gardner, A., West, S. A., & Barton, N. H. (2007). The Relation between Multilocus Population Genetics and Social Evolution Theory. *The American Naturalist*. https://doi.org/10.1086/510602

Gazzillo, S., Sopher, B., Aktipis, A., & Cronk, L. (2013). The origins of risk sharing: An experimental approach. *Working Paper No. 2013-02*.

Ghang, W., & Nowak, M. A. (2014). Stochastic evolution of staying together. *Journal of Theoretical Biology*. https://doi.org/10.1016/j.jtbi.2014.06.026

Ghang, W., & Nowak, M. A. (2015). Indirect reciprocity with optional interactions. *Journal of Theoretical Biology*. https://doi.org/10.1016/j.jtbi.2014.09.036

Gilbert, O. M., Foster, K. R., Mehdiabadi, N. J., Strassmann, J. E., & Queller, D. C. (2007). High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(21), 8913–8917.

Grafen, A. (1985). A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology*, *2*(2).

Grafen, A. (2006). Optimization of inclusive fitness. *Journal of Theoretical Biology*. https://doi.org/10.1016/j.jtbi.2005.06.009

Grafen, A. (2007). An inclusive fitness analysis of altruism on a cyclical network.

Journal of Evolutionary Biology, *20*(6), 2278–2283. https://doi.org/10.1111/j.1420-9101.2007.01413.x

Grafen, A. (2015). Biological fitness and the fundamental theorem of natural selection. *The American Naturalist*, *186*(1), 1–14.

Hamilton, W. D. (1964). The genetical theory of social behavior. I and II. *Journal of Theoretical Biology*, *7*(1), 1–52.

Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature*, *228*(5277), 1218–1220.

Hao, Y., Armbruster, D., Cronk, L., & Aktipis, C. A. (2015). Need-based transfers on a network: A model of risk-pooling in ecologically volatile environments. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2014.12.003

Hauser, O. P., Nowak, M. A., & Rand, D. G. (2014). Punishment does not promote cooperation under exploration dynamics when anti-social punishment is possible. *Journal of Theoretical Biology*. https://doi.org/10.1016/j.jtbi.2014.06.041

Hauser, O. P., Rand, D. G., Peysakhovich, A., & Nowak, M. A. (2014). Cooperating with the future. *Nature*. https://doi.org/10.1038/nature13530

Hecht, E. E., Robins, D. L., Gautam, P., & King, T. Z. (2017). Intranasal oxytocin reduces social perception in women: Neural activation and individual variation. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2016.12.046

Hilbe, C., Hoffman, M., & Nowak, M. (2015). Cooperate without Looking in a Non-Repeated Game. *Games*. https://doi.org/10.3390/g6040458

Hilbe, C., Nowak, M. A., & Sigmund, K. (2013). Evolution of extortion in Iterated Prisoner's Dilemma games. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1214834110

Hilbe, C., Nowak, M. A., & Traulsen, A. (2013). Adaptive dynamics of extortion and compliance. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0077886

Hilbe, C., Wu, B., Traulsen, A., & Nowak, M. A. (2014). Cooperation and control in multiplayer social dilemmas. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1407887111

Hilbe, C., Wu, B., Traulsen, A., & Nowak, M. A. (2015). Evolutionary performance of zero-determinant strategies in multiplayer games. *Journal of Theoretical Biology*. https://doi.org/10.1016/j.jtbi.2015.03.032

Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: Why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1417904112

Hughes, W. O. H., Oldroyd, B. P., Beekman, M., & Ratnieks, F. L. W. (2008). Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science*, *320*(5880), 1213–1216.

Inglis, R. F., Ryu, E., Asikhia, O., Strassmann, J. E., & Queller, D. C. (2017). Does high relatedness promote cheater-free multicellularity in synthetic lifecycles? *Journal of Evolutionary Biology*. https://doi.org/10.1111/jeb.13067

Jack, C. N., Adu-Oppong, B., Powers, M., Queller, D. C., & Strassmann, J. E. (2011). Cost of movement in the multicellular stage of the social amoebae Dictyostelium discoideum and D. purpureum. *Ethology Ecology and Evolution*. https://doi.org/10.1080/03949370.2011.584907

Jack, C. N., Buttery, N., Adu-Oppong, B., Powers, M., Thompson, C. R. L., Queller, D. C., & Strassmann, J. E. (2015). Migration in the social stage of *Dictyostelium discoideum* amoebae impacts competition. *PeerJ*.

https://doi.org/10.7717/peerj.1352

Jordan, J. J., Sommers, R., Bloom, P., & Rand, D. G. (2017). Why Do We Hate Hypocrites? Evidence for a Theory of False Signaling. *Psychological Science*. https://doi.org/10.1177/0956797616685771

Jordan, J., McAuliffe, K., & Rand, D. (2016). The effects of endowment size and strategy method on third party punishment. *Experimental Economics*. https://doi.org/10.1007/s10683-015-9466-8

Kraft-Todd, G., Yoeli, E., Bhanot, S., & Rand, D. (2015). Promoting cooperation in the field. *Current Opinion in Behavioral Sciences*. https://doi.org/10.1016/j.cobeha.2015.02.006

Krasnow, M. M., Cosmides, L., Pedersen, E. J., & Tooby, J. (2012). What Are Punishment and Reputation for? *PLoS ONE*. https://doi.org/10.1371/journal.pone.0045662

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2015). Group cooperation without group selection: Modest punishment can recruit much cooperation. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0124561

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking Under the Hood of Third-Party Punishment Reveals Design for Personal Benefit. *Psychological Science*, *27*(3), 405–418. https://doi.org/10.1177/0956797615624469

Krasnow, M. M., Delton, A. W., Tooby, J., & Cosmides, L. (2013). Meeting now suggests we will meet again: Implications for debates on the evolution of cooperation. *Scientific Reports*. https://doi.org/10.1038/srep01747

Krebs, J. R., & Davies, N. B. (1978). *Behavioural ecology: an evolutionary approach*. John Wiley & Sons.

Krebs, J. R., & Davies, N. B. (1987). An introduction to behavioral ecology. 2nd. *John Wiley & Sons*.

Krebs, J. R., & Davies, N. B. (2009). *Behavioural ecology: an evolutionary approach*. John Wiley & Sons.

Kurzban, R., Burton-chellew, M. N., & West, S. A. (2015). The Evolution of Altruism in Humans. https://doi.org/10.1146/annurev-psych-010814-015355

Kuzdzal-Fick, J. J., Fox, S. A., Strassmann, J. E., & Queller, D. C. (2011). High relatedness is necessary and sufficient to maintain multicellularity in Dictyostelium. *Science*, *334*(6062), 1548–1551.

Laland, K. (2017). *Darwin's unfinished symphony: How culture made the human mind*.

Laland, K. N. (2017). The Foundations of Human Cooperation in Teaching and Imitation. *Spanish Journal of Psychology*. https://doi.org/10.1017/sjp.2016.101

Leverett, K., Heyler, C., Flemming, T., Talbot, C. F., Zak, P. J., Essler, J. L., … Brosnan, S. F. (2015). Oxytocin reduces food sharing in capuchin monkeys by modulating social distance. *Behaviour*. https://doi.org/10.1163/1568539X-00003268

Levin, S. R., Brock, D. A., Queller, D. C., & Strassmann, J. E. (2015). Concurrent coevolution of intra-organismal cheaters and resisters. *Journal of Evolutionary Biology*. https://doi.org/10.1111/jeb.12618

Liao, X., Rong, S., & Queller, D. C. (2015). Relatedness, Conflict, and the Evolution of Eusociality. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.1002098

Manapat, M. L., Nowak, M. A., & Rand, D. G. (2013). Information, irrationality, and the evolution of trust. *Journal of Economic Behavior and Organization*. https://doi.org/10.1016/j.jebo.2012.10.018

Nielsen, M., Mushin, I., Tomaselli, K., & Whiten, A. (2016). Imitation, Collaboration, and Their Interaction Among Western and Indigenous Australian Preschool Children. *Child Development*. https://doi.org/10.1111/cdev.12504

Nishi, A., Christakis, N. A., Evans, A. M., O'Malley, A. J., & Rand, D. G. (2016). Social Environment Shapes the Speed of Cooperation. *Scientific Reports*. https://doi.org/10.1038/srep29622

Nishi, A., Shirado, H., Rand, D. G., & Christakis, N. A. (2015). Inequality and visibility of wealth in experimental social networks. *Nature*. https://doi.org/10.1038/nature15392

Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature*, *466*(7310), 1057–1062.

Ohtsuki, H., Hauert, C., Lieberman, E., & Nowak, M. A. (2006). *A simple rule for the evolution of cooperation on graphs*. *Nature* (Vol. 441).

Ohtsuki, H., Iwasa, Y., & Nowak, M. A. (2015). Reputation Effects in Public and Private Interactions. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1004527

Olejarz, J., Ghang, W., & Nowak, M. A. (2015). Indirect Reciprocity with Optional Interactions and Private Information. *Games*, *6*, 438–457.

Olejarz, J. W., & Nowak, M. A. (2014). Evolution of staying together in the context of diffusible public goods. *Journal of Theoretical Biology*. https://doi.org/10.1016/j.jtbi.2014.06.023

Ostrowski, E. A., Shen, Y., Tian, X., Sucgang, R., Jiang, H., Qu, J., … Queller, D. C. (2015). Genomic Signatures of Cooperation and Conflict in the Social Amoeba. *Current Biology*. https://doi.org/10.1016/j.cub.2015.04.059

Petersen, M. B., Sell, A., Tooby, J., & Cosmides, L. (2012). To punish or repair? Evolutionary psychology and lay intuitions about modern criminal justice. *Evolution and Human Behavior*, *33*(6), 682–695. https://doi.org/10.1016/j.evolhumbehav.2012.05.003

Petersen, M. B., Sznycer, D., Cosmides, L., & Tooby, J. (2012). Who Deserves Help? Evolutionary Psychology, Social Emotions, and Public Opinion about Welfare. *Political Psychology*. https://doi.org/10.1111/j.1467-9221.2012.00883.x

Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a "cooperative phenotype" that is domain general and temporally stable. *Nature Communications*. https://doi.org/10.1038/ncomms5939

Pietraszewski, D., Cosmides, L., & Tooby, J. (2014). The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0088534

Pietraszewski, D., Curry, O. S., Petersen, M. B., Cosmides, L., & Tooby, J. (2015). Constituents of political cognition: Race, party politics, and the alliance detection system. *Cognition*. https://doi.org/10.1016/j.cognition.2015.03.007

Queller, D. C. (1985). Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature*, *318*(6044), 366–367.

Queller, D. C. (1992). A general model for kin selection. *Evolution*, *46*(2), 376–380.

Queller, D. C. (1994). Genetic relatedness in viscous populations. *Evolutionary Ecology*, *8*(1), 70–73.

Queller, D. C. (1997). Cooperators Since Life Began. *The Quarterly Review of Biology*, *72*(2), 184–188. Retrieved from http://www.jstor.org/stable/3036338

Queller, D. C. (2014). Joint phenotypes, evolutionary conflict and the fundamental theorem of natural selection. *Philosophical Transactions of the Royal Society B:*

Biological Sciences. https://doi.org/10.1098/rstb.2013.0423

Queller, D. C. (2016a). Kin Selection and Its Discontents. *Philosophy of Science*. https://doi.org/10.1086/687870

Queller, D. C. (2016b). *Theory of Inclusive Fitness*. Retrieved from http://openscholarship.wustl.edu/bio_facpubshttp://openscholarship.wustl.edu/bio_facpubs/121

Queller, D. C. (2017). Fundamental Theorems of Evolution. *The American Naturalist*. https://doi.org/10.1086/690937

Queller, D. C., Rong, S., & Liao, X. (2015). Some Agreement on Kin Selection and Eusociality? *PLoS Biology*. https://doi.org/10.1371/journal.pbio.1002133

Queller, D. C., & Strassmann, J. E. (n.d.). *III.4 Kin Selection and Inclusive Fitness*.

Queller, D. C., & Strassmann, J. E. (2013). Experimental evolution of multicellularity using microbial pseudo-organisms. *Biology Letters*. https://doi.org/10.1098/rsbl.2012.0636

Queller, D. C., & Strassmann, J. E. (2016). Problems of multi-species organisms: endosymbionts to holobionts. *Biology and Philosophy*. https://doi.org/10.1007/s10539-016-9547-x

Rand, D. G. (2016). Cooperation, Fast and Slow: Meta-Analytic Evidence for a Theory of Social Heuristics and Self-Interested Deliberation. *Psychological Science*. https://doi.org/10.1177/0956797616654455

Rand, D. G., Brescoll, V. L., Everett, J. A. C., Capraro, V., & Barcelo, H. (2016). Social Heuristics and Social Roles: Intuition Favors Altruism for Women but Not for Men. https://doi.org/10.1037/xge0000154.supp

Rand, D. G., Dreber, A., Haque, O. S., Kane, R. J., Nowak, M. A., & Coakley, S. (2014). Religious motivations for cooperation: An experimental investigation using explicit primes. *Religion, Brain and Behavior*. https://doi.org/10.1080/2153599X.2013.775664

Rand, D. G., Fudenberg, D., & Dreber, A. (2015). It's the thought that counts: The role of intentions in noisy repeated games. *Journal of Economic Behavior and Organization*. https://doi.org/10.1016/j.jebo.2015.05.013

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. https://doi.org/10.1038/nature11467

Rand, D. G., Nowak, M. A., Fowler, J. H., & Christakis, N. A. (2014). Static network structure can stabilize human cooperation. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1400406111

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*. https://doi.org/10.1038/ncomms4677

Rand, D. G., Tarnita, C. E., Ohtsuki, H., & Nowak, M. A. (2013). Evolution of fairness in the one-shot anonymous Ultimatum Game. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1214167110

Reiter, J. G., Hilbe, C., Rand, D. G., Chatterjee, K., & Nowak, M. A. (2018). Crosstalk in concurrent repeated games impedes direct reciprocity and requires stronger levels of forgiveness. *Nature Communications*. https://doi.org/10.1038/s41467-017-02721-8

Robertson, T. E., Delton, A. W., Klein, S. B., Cosmides, L., & Tooby, J. (2014). Keeping the benefits of group cooperation: Domain-specific responses to distinct causes of social exclusion. *Evolution and Human Behavior*. https://doi.org/10.1016/j.evolhumbehav.2014.06.006

Rogers, C. N., Ross, A. P., Sahu, S. P., Siegel, E. R., Dooyema, J. M., Cree, M. A.,

… Preuss, T. M. (2018). Oxytocin- and arginine vasopressin-containing fibers in the cortex of humans, chimpanzees, and rhesus macaques. *American Journal of Primatology*. https://doi.org/10.1002/ajp.22875

Roithmayr, D., Isakov, A., & Rand, D. G. (2014). *Should Law Keep Pace with Society? Relative Update Rates Determine the Co-Evolution of Institutional Punishment and Citizen Contributions to Public Goods*. SSRN. https://doi.org/10.2139/ssrn.2404931

Shu, L., Zhang, B., Queller, D. C., & Strassmann, J. E. (2018). Burkholderia bacteria use chemotaxis to find social amoeba Dictyostelium discoideum hosts. *ISME Journal*. https://doi.org/10.1038/s41396-018-0147-4

Smith, J. M., & Szathmary, E. (1995). *The major transitions in evolution*. Oxford University Press.

Smith, J., Queller, D. C., & Strassmann, J. E. (2014). Fruiting bodies of the social amoeba Dictyostelium discoideum increase spore transport by Drosophila. *BMC Evolutionary Biology*. https://doi.org/10.1186/1471-2148-14-105

Smith, J., Strassmann, J. E., & Queller, D. C. (2016). Fine-scale spatial ecology drives kin selection relatedness among cooperating amoebae. *Evolution*. https://doi.org/10.1111/evo.12895

Stagnaro, M. N., Arechar, A. A., & Rand, D. G. (2017). From good institutions to generous citizens: Top-down incentives to cooperate promote subsequent prosociality but not norm enforcement. *Cognition*. https://doi.org/10.1016/j.cognition.2017.01.017

Strassmann, J. E. (2016). Kin Discrimination in Dictyostelium Social Amoebae 1, 378–383. https://doi.org/10.1111/jeu.12307

Strassmann, J. E., & Queller, D. C. (n.d.). *VII.9 Cooperation and Conflict: Microbes to Humans*. Retrieved from http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=1372264.

Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S., & Halperin, E. (2016). Shame closely tracks the threat of devaluation by others, even across cultures. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1514699113

Tarnita, C. E., Taubes, C. H., & Nowak, M. A. (2013). Evolutionary construction by staying together and coming together. *Journal of Theoretical Biology*. https://doi.org/10.1016/j.jtbi.2012.11.022

Taylor, P. (2017). Inclusive fitness in finite populations – effects of heterogeneity and synergy. *Evolution*, *71*(3), 508–525.

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., … Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*. https://doi.org/10.1038/nature12194

van Gestel, J., & Tarnita, C. E. (2017). On the origin of biological construction, with a focus on multicellularity. *Proceedings of the National Academy of Sciences*, *114*(42), 11018–11026. https://doi.org/10.1073/pnas.1704631114

West, S. A., Fisher, R. M., Gardner, A., & Kiers, E. T. (2015). Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences*, *112*(33), 10112–10119. https://doi.org/10.1073/pnas.1421402112

West, S. A., & Gardner, A. (2013). Adaptation and inclusive fitness. *Current Biology*, *23*(13), R577–R584.

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, *20*(2), 415–432.

Westneat, D., & Fox, C. W. (2010). *Evolutionary behavioral ecology*. Oxford University Press.

Yoeli, E., Hoffman, M., Rand, D. G., & Nowak, M. A. (2013). Powering up with indirect reciprocity in a large-scale field experiment. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1301210110