

Empathy, compassion and the humanity paradox: a transjective account

Paper II of the Peace Papers, six papers on one framework; each stands alone. Website version, working draft v18, 30 October 2026. Devon Summerhayes, London.

Abstract

Scientific and spiritual accounts of human experience are usually treated as rivals. This essay treats them as two views of one experience, which it calls transjective, and applies the view to empathy and compassion. It makes three contributions. First, a labelling discipline: every descriptive claim is marked Experienced, Measured or Interpretive, so that the two languages do not undermine each other. Second, an arrangement of three perspectives of the I: a felt perspective (I1, emotional empathy, with a positive pole, a negative pole in two forms, reactive and appetitive, and a neutral setting), an understanding perspective (I2, cognitive empathy) and the witness's eye (I3), through which the other two are seen together. Compassion is felt concern directed by understanding and acted on. Crossing the three settings of I1 with the three roles of I2 gives nine combinations; five of them hold six named states, compassion, indifference, defended detachment, cruelty, revenge and distress, since cruelty and revenge share one combination and differ by the form of I1-. Three of them, indifference, cruelty and revenge, are states the usual contrast of empathic concern with personal distress does not separate. The arrangement is applied to the humanity paradox: the capacity that lets humans cooperate at scale, understanding other minds, is the capacity that lets a few exploit the rest, and cooperation itself creates the niche in which exploitation pays. Game theory shows why cooperation is hard and why it happens all the same. The essay's answer has two halves and is Interpretive: in the person, compassion, the two empathies rejoined so that understanding serves others; in the population, structures such as repeated dealing, visible reputation, monitoring and fair rules, which contain the few without relying on felt concern. Third, four testable hypotheses are stated, each with what would count against it. The framework comes from the author's introspection and is offered for testing, not as a finding.

In the series, the paper takes its perspectives from paper I and supplies the states and the humanity paradox that the later papers use: paper III reads the paradox's puzzle as two perspectives taken for one, paper IV carries revenge and forgiveness into the loops that pass between generations, and paper VI finds the same paradox at scale in artificial intelligence.

1. Introduction

Two accounts of a person are usually kept apart. One measures: how biology and environment shape emotion and behaviour. The other feels: the same life described in the words of spirit, conscience and soul. Side by side without a rule for telling them apart, each discredits the other. This essay keeps both and adds the rule: every descriptive claim says which kind of claim it is, and a third stance, the witness's, holds the two accounts together without dismissing either. The companion essay, "Arriving at the centre of consciousness: a transjective framework" (paper I; Summerhayes, 2026, under review), sets out that stance (transjectuality) and the practice of working from it (scientuality) in full; this essay takes from it only what the argument needs and applies it to empathy, compassion and the humanity paradox.

The perspectives, in brief (author's mapping, Interpretive; in full in the companion essay, paper I, section 1). The I is the person's first-person self. Three perspectives are taken from it, each a place experience is taken up from, not a part of the mind or a type of person. I1 is the felt perspective, subjective: the inner life as it is lived, which this essay ties to emotional empathy and calls spiritual in the sense of inner experience, not a separate substance. It runs as one range from a negative pole, I1-, through a neutral setting to a positive pole, I1+; the range is an Interpretive picture, and the settings are measured as separate variables (Table 5). At I1+ are conscience, trust,

grace and felt connection; at neutral, calm, unengaged feeling; at I1-, the self-directed pole, in two forms, reactive (fear, hurt and anger, being overwhelmed) and appetitive (appetite, instrumental aggression), with felt concern for others absent in both. I2 is the understanding perspective, objective in the sense that it takes experience as its object: analytic thought, tied here to cognitive empathy, neutral in itself and able to serve any setting of I1. It plays one of three roles: offline (I1 “alone”), present while feeling sets the course (I1 “with” I2), or setting the course (I1 “directed by” I2); a person or a system “directed by” a feeling is the ordinary sense, the feeling that sets the aim. I3 is the witness’s eye, through which feeling and understanding are seen together; the witness is the self looking through it, and transjectuality the stance it has; the eye does not act, and the witness’s part is to notice which state one is in, which makes moving between states possible. A setting of I1 is the presence and direction of felt concern (present, absent, or turned on the self alone), and a form is the drive within a setting, so the two forms of I1- are one setting: three settings crossed with three roles of I2 give nine combinations, of which Table 3 names six as states: compassion (I1+ directed by I2), indifference (neutral I1 with I2), defended detachment (I1- reactive with I2), cruelty (I1- appetitive directed by I2), revenge (I1- reactive directed by I2) and distress (I1- reactive alone). The states of I1- are told apart by form within their setting, which is why six named states and the unnamed combinations do not sum to nine. Figure 1 draws the three perspectives; Figure 2 draws compassion as the seam of the two empathies, the join where the two meet and are held together, seen from I3.

Three perspectives of the I: a triangle in its circle

I1- and I1+ are the poles of the base, I2 is the apex, I3 is the centre of the circle (Interpretive)

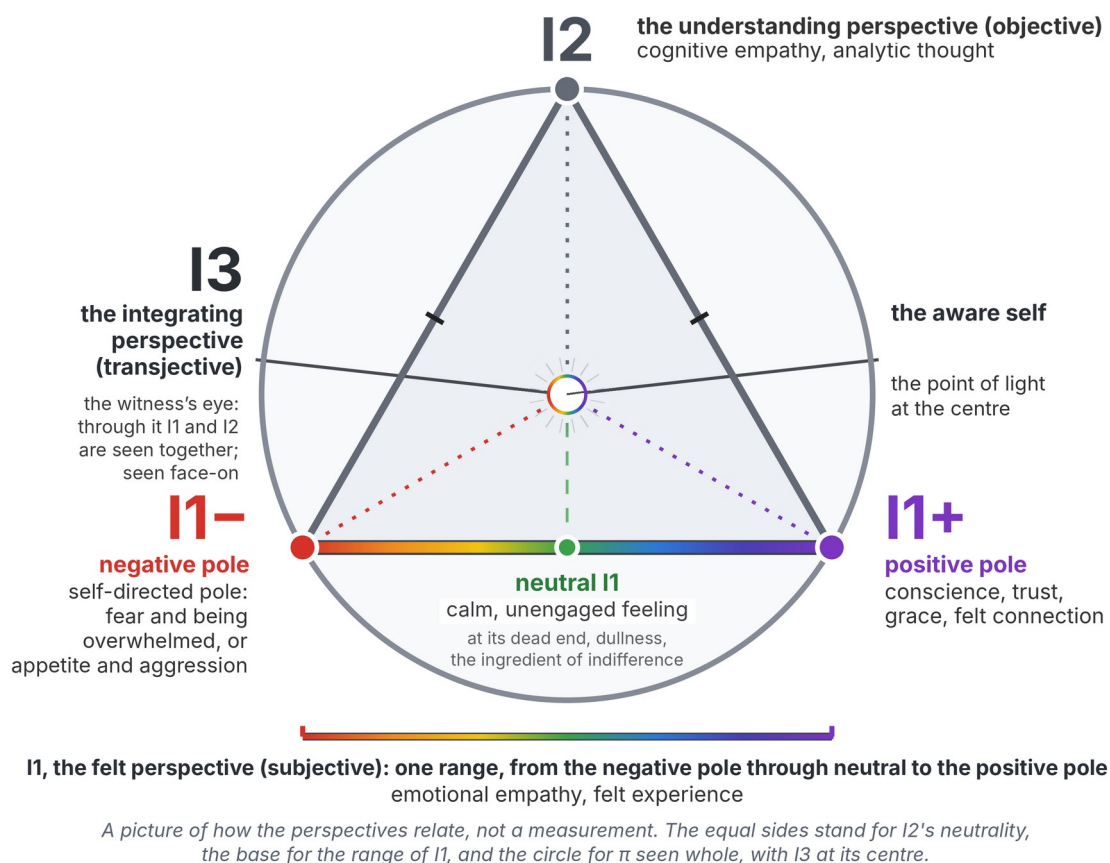


Figure 1. The perspectives of the I drawn as a triangle in its circle (Interpretive). I1- and I1+ are the poles of the felt perspective at the ends of the base, neutral I1 between them; I2, the understanding perspective, is the apex; I3, the witness's eye, is the ring at the centre, with the aware self as the point of light within it. Emotional empathy sits at I1 and cognitive empathy at I2; compassion combines them (Table 3) and has no position on the figure. A picture of how the perspectives relate, not a measurement.

Position of the author

The author is an independent researcher, and the framework began as self-examination and informal observation of other people over many years, checked afterwards against published research. Both sources are limited: people misjudge their own minds and other people's, and unsystematic observation invites bias. The framework is offered as a hypothesis for others to test. No identifiable individual is described and no case material is reported. Measured claims come from published sources and are marked; Experienced and Interpretive claims are the author's own. A scheme like this is a map, and a map can become a hiding place, a way of sorting life into categories instead of meeting its harder parts; the three modes are meant to keep the felt account in view beside the measured one. The author founded a community project, named in the Declarations, whose aims relate to the practical side of this essay.

How the framework was developed and checked. The scheme was drafted from the author's observations and revised as inconsistencies surfaced. There was no systematic literature search: sources were chosen to test and sharpen the scheme, so they may lean towards support, and were not appraised for quality. Numbers were checked against the original papers or their abstracts. Generative AI helped with drafting, consistency checks and reference checks (Anthropic, 2026; see Declarations).

2. What "transjective" means here

Transjective names what is neither purely subjective nor purely objective: meaning that arises between a knower and what is known, as the shade of a tree is in neither the tree alone nor the person standing in it. The word is rare; John Vervaeke uses it for the relation between agent and world (Vervaeke, 2019), and it is used here in that plain sense, which the series' paradox paper sharpens into a stricter sense: the subjective and the objective held as one from the stance that sees both; there is no claim to have coined it. The empathy spectrum shows why the word is needed. A scientist sees a trait that varies across people and is measured by questionnaire and by the brain's response to another's pain (Davis, 1983; Singer et al., 2004); another person sees conscience, heart or soul. These are not rival claims but one thing described from outside and from inside (Interpretive).

Related work and what is new (Measured and Interpretive)

The words "transjectuality", for the stance taken through I3, and "scientuality", for the practice of holding the two languages together, are the author's own and are introduced in the companion essay (paper I, section 2), which also discusses the literatures nearest the approach: typologies of science and religion, neurophenomenology, dual-aspect monism, decentering and the minimal self. Here the witness is taken as given, and the nearest relatives are in empathy research.

The felt and understanding perspectives have close relatives in psychology. Empathy research separates affective from cognitive components (Decety & Jackson, 2004; Zaki & Ochsner, 2012) and traces them to perception-action mechanisms with deep evolutionary roots (Preston & de Waal, 2002). Batson (2011) links empathic concern to helping, and Goetz et al. (2010) define compassion as concern for another's suffering with a wish to help, distinct from empathy, as Singer and Klimecki (2014) do. In this essay compassion means the act: felt concern (I1+), directed by understanding (I2), and acted on to help (Table 3). The concern with a wish to help that those authors call compassion is its felt part, and without the act it is called concern here. Bloom (2016) argues that felt empathy is a poor guide to moral action and that reasoned compassion serves better; the cognitive rival to H1 in section 7 is in that spirit. People often avoid empathy when they expect it to be effortful (Cameron et al., 2019), so an avoidance measure in H4 may reflect effort as well as self-protection.

Training studies bear on the same distinction. In one, empathy training increased negative affect, whereas compassion training reversed it and engaged a different brain network (Klimecki et al., 2014); in another, two weeks of compassion training, compared with reappraisal training, increased altruistic spending on a victim (Weng et al., 2013). In this essay's terms the first pattern resembles

the felt side of the distress row of Table 3 and the second the compassion row. The samples are small and the mapping is the author's, so these studies motivate H1 and H4 and do not test them. The scheme also overlaps with an older distinction: Batson (2011) separates empathic concern, which is other-oriented, from personal distress, which is self-oriented. Here I1+ resembles the first and the reactive form of I1-, overwhelmed, resembles the second, so those two settings are not new. This essay adds no new mechanism; its additions are the labelling discipline, the arrangement of Table 3, whose named states include three, indifference, cruelty and revenge, that the concern-distress contrast does not separate, and four hypotheses.

3. The discipline: three modes of claim

Each section or paragraph of this essay marks the mode of its descriptive claims, and a sentence carries its own label where it departs from that mode; Table 1 sets out the three, and the companion essay states the rule in full (paper I, section 3). Definitions (what a word means here) and statements of value or advice are not claims about the world, so they sit outside the modes and are marked as such where they occur. Two consequences follow: words like "proves" belong only to Measured claims, and physics and mathematics may be used as images in Experienced or Interpretive claims, but not as evidence for souls.

Table 1. The three modes of claim.

Mode	Example	How to judge it
Experienced	A tightness in the chest on hearing of a friend's loss, and a wish to sit with her.	Honesty, and consistency with other accounts: testimony, not proof, and not to be defended as physics.
Measured	Two weeks of compassion training increased altruistic redistribution in a laboratory game, compared with training in reappraisal (Weng et al., 2013).	Evidence, replication and effect size.
Interpretive	Compassion is felt concern, directed by understanding, and acted on to help.	Coherence and usefulness: whether it clarifies more than it distorts.

A worked example

The companion essay describes one moment in three modes (paper I, section 3): a friend says that her father has died; the tightness in the chest is Experienced, the shared affective response to another person's distress is Measured (Singer et al., 2004), and reading the tightness as caring is Interpretive.

The same moment can also go four ways for the person who hears the news, according to the setting of I1 and the role of I2 (Table 3; the cruelty and revenge rows are left out here):

- **Compassion (I1+ directed by I2):** you sit with her, listen, and offer the help you have judged she needs.
- **Indifference (neutral I1 with I2):** you understand what has happened and say the right words, but you are not moved.
- **Defended detachment (I1-, reactive form, with I2):** you understand, but the feeling is too much to hold and sets the aim, which is escape, so you change the subject or find a reason to leave.
- **Distress (I1-, reactive form, alone):** you are overwhelmed by the news yourself and need comfort, and you are no help to her.

4. The humanity paradox, and the mismatch that sharpens it (Measured and Interpretive)

The “humanity paradox”, as used here, is this: the capacity that lets humans cooperate at scale, understanding other minds, is the same capacity that lets a few exploit the rest, and cooperation itself creates the niche in which exploitation pays. The mechanism is that emotional empathy and cognitive empathy come apart. People differ widely in felt concern for others (Measured as individual differences on the Empathic Concern scale; Davis, 1983), with concern reduced at the low end in the clinical groups of section 6, while the ability to understand what another person thinks and feels is dissociable from it: lesion work shows the two come apart (Shamay-Tsoory et al., 2009), and in psychopathic traits emotional empathy is reduced while cognitive empathy is relatively preserved (Shukla & Upadhyay, 2025; section 6). So understanding can serve others or the self, depending on the feeling behind it, and the more a population cooperates, the more there is to gain for the few whose understanding serves feeling at its negative pole (section 5). That is the paradox: the capacity that lets humans cooperate is the capacity that lets a few exploit the rest, and the better they cooperate, the more exploitation pays. The dual-use point is not new. The Machiavellian-intelligence hypothesis holds that social intellect evolved in part for manipulation (Byrne & Whiten, 1988; Humphrey, 1976), and recent work describes a “dark empath” profile, dark traits with empathy, cognitive and affective, preserved (Heym et al., 2021). What the framework adds is the placement: understanding is I2, neutral in itself, and the setting of I1 it serves decides the state (Table 3), so kindness is not something understanding brings or fails to bring but the whole range of I1 beside it, from callous to warm, with understanding serving any point of that range. The paradox rests on no assumption that understanding makes us kind; it is in the world, and it is a paradox of peace: the capacity peace needs is the capacity that undoes it, and the better a people cooperates, the more it pays not to. Seen as two, a neutral capacity and the feeling it serves, the puzzle of one gift with opposite uses goes (paper III, section 1), and what remains, dual use and the niche it creates, is the tension the rest of the essay answers in two halves (section 9), since no seeing removes it. Some fraction of any generation can be expected to put understanding in the service of appetite, by partly heritable disposition (callous-unemotional traits showed strong genetic influence in a large twin sample of seven-year-olds, teacher-rated; Viding et al., 2005; that the disposition persists into adult use is a further step), by development or by incentive: the cruelty row of Table 3 and the harm profile of H4. A second route runs through injury rather than appetite (Interpretive; its measures are in Table 5): the reactive form of I1– directed by I2, hurt made into a plan, which Table 3 names revenge; paper IV (section 3.1) follows it as a loop that passes between generations. Revenge answers an injury to identity (identity and difference are read in section 6: concern knows the other as not other, understanding knows the other as other), being made other by those who should have known one as the same, and it demands identity back by force, “now you know what I felt”; the demand cannot be met by that route, since the other’s suffering stays the other’s, so each blow asks for another, and the cycle rebuilds itself in the way a group that makes others of its own members cannot close by one more sanction. Its opposite is forgiveness: the same injury met from I1+, concern directed by understanding toward the one who caused it, which is compassion’s own combination under injury; the move from revenge to forgiveness is a change in the setting of I1, along the base of Figure 1 from the negative pole to the positive, with understanding still directing, and it is made from I3, where the loop is seen as a loop. Between them sits the grudge: the injury held with understanding and the feeling walled off, defended detachment under injury, which can tip either way; once the feeling has gone cold, a grievance kept but no longer felt, it sits at neutral (Interpretive). That is the humanity paradox from the other side: the fraction who take revenge is partly made, by exclusion, humiliation or use, and not only born. A peace that depends on everyone’s goodwill is therefore out of reach, and the question becomes what lets the many cooperate when the few do not. Section 5 shows why such a fraction can be stable, in models, and what may enlarge it; the term says nothing about any individual (Interpretive). The capacity is now also a technology: a system that models other minds at scale is understanding without a feeling of its own, and the paradox applies to it unchanged, with the setting of I1 supplied by whoever directs it (Interpretive).

The mismatch that sharpens it (Interpretive, with the biology Measured). One reason the paradox bites harder now is a mismatch between the biology we inherited and the world we built (Gluckman et al., 2009; Nesse & Williams, 1994). The conditions that once made cooperation pay, repeated contact, visible reputation and shared fate, are weaker in large, anonymous, fast societies, so the fraction that plays for itself costs more and is harder to contain (section 5). The claim is not that the past was better: human bodies have kept evolving, there is no single ancestral environment to return to, and modern hunter-gatherers are not models of early humans (Zuk, 2013). The claim is only that some modern conditions sit awkwardly with our inheritance, and section 5 says why that matters for cooperation; Measured work can test only narrow pieces of it, and the wider reading is Interpretive.

5. Game theory: why cooperation is hard, and why it happens (Measured and Interpretive)

Why this section: if understanding can serve the self as readily as others, some fraction of any population can be expected to use it that way, and the humanity paradox then asks what lets the many cooperate when the few do not. Game theory is the formal language for that question, and for the related one of why choices that are sensible for each person can add up to a worse outcome; it also shows where the felt perspective and the understanding perspective each enter.

Game theory studies choices whose results depend on what others choose.

- **The dilemma and the commons.** In a one-off prisoner's dilemma, defecting pays each person even though mutual cooperation is better for all. Hardin's "tragedy of the commons" (1968) has the same shape. When interactions repeat, cooperation through reciprocity such as tit-for-tat can be stable (Axelrod, 1984). Tit-for-tat is revenge formalised, defection answered with defection, and where mistakes happen it locks two players into endless retaliation, the self-rebuilding cycle of section 4; what does better there is generous tit-for-tat, which forgives a share of defections and so breaks the loop (Nowak & Sigmund, 1992; the same release ends the vendetta in paper III, section 3, Table 1, and the revenge form of cyclic harm in paper IV, section 3.1), forgiveness favoured in simulated populations, though later models show the cooperating population can itself give way once unconditional cooperators drift in and defectors return, a cycle of cooperation and defection (Imhof et al., 2005) (results in models, not in people; the reading Interpretive).
- **Communities often avoid the tragedy.** Elinor Ostrom (1990) found that long-lasting commons share design features, among them clear boundaries, local rules, inclusive decisions, monitoring by users, graduated penalties and recognised self-governance.
- **Routes to cooperation.** Nowak (2006) lists five: kin selection, direct reciprocity, indirect reciprocity (reputation), network reciprocity and group selection.
- **Hawk-Dove.** Maynard Smith and Price (1973) introduced the evolutionarily stable strategy, and the Hawk-Dove game showed that when aggression pays only while rare, a population can settle into a stable mix of aggressive and peaceful strategies (Maynard Smith, 1982). Mealey (1995) applied similar logic to sociopathy, treating it as a frequency-dependent life strategy with two developmentally different routes; it was criticised (for example Crusio, 1995) and remains disputed. This is the population-level shape of the paradox of section 4: a fraction that puts understanding in the service of appetite can be stable, not because anyone is a devil, but because the strategy pays while it is rare. It must never be used to label individuals.
- **Mismatch as changed payoffs (Interpretive).** Small groups with repeated contact, visible reputations and shared fate make cooperation pay. Large, anonymous, fast societies look more like one-off games: reputations are harder to track and defection is cheaper. The evidence cuts both ways: across fifteen small-scale societies, people more integrated into

markets made fairer offers to strangers, not less fair (Henrich et al., 2001, 2010), and large societies build reputation systems of their own, credit records, reviews and courts, so anonymity is not the whole story. The hypothesis is therefore narrower than it sounds: that our biology evolved for repeated play, that the world increasingly offers one-shot play as well, and that the balance of incentives decides which fraction grows. If the fraction that plays for itself does not grow where one-shot incentives are introduced, or shrink where repeated dealing is restored, this reading fails. Incentives that reward one-shot play can enlarge the fraction that plays for itself, and those that reward repeated dealing can shrink it, inside markets and outside them; the claim is about incentives, not any one economic system (Interpretive).

- **Beyond payoffs.** People do not only maximise payoffs. Emotions such as guilt, loyalty and compassion, and also anger and the urge to retaliate, can act as commitment devices (Frank, 1988), and costly shared practices can signal commitment: in a study of 83 nineteenth-century US communes, religious communes with costlier requirements lasted longer, and secular communes did not show the effect (Sosis & Bressler, 2003). Experiments in small-scale societies also show people cooperating more than pure self-interest predicts, with wide variation between cultures (Henrich et al., 2001).

Game-theory models simplify. They help most as a way of asking what would make cooperation pay, not as a prediction of how any person will act (Interpretive).

Systems as understanding made durable (Interpretive). The designs just named are I2 written down: rules, measures, roles and incentives that keep working when no one is feeling anything, which is their strength and their failure in one. A system carries on without I1, which is what lets a hospital run at three in the morning, and the same absence is what lets an office harm a person by following the procedure. A system also freezes the feeling it was built with: a hospice, a credit union or a mutual-aid network is an I2 structure with I1+ set into it, and it keeps caring after its founders have gone home, which is the one thing feeling cannot do for itself. So the practical question is not less I2 but I2 built from I1+ and still witnessed: structures that build in a return to the felt perspective and a place for someone whose role is to look rather than to run. This design claim is not tested here.

6. Compassion: where the two empathies meet (Interpretive)

Sections 4 and 5 set the problem: the two empathies come apart, so understanding can serve the self as readily as others, in a world mismatched to our inheritance where the conditions that once made cooperation pay, repeated contact, visible reputation and shared fate, are weaker. Section 5 gave two answers, surroundings designed so that cooperating is the stable choice, and the felt commitments that change the game from the inside. This section is about the second. Compassion, felt concern directed by understanding and acted on, is the state in which the two empathies are rejoined and do together what neither does alone. It is a marriage of the two empathies within one person: each keeps its own nature, they are joined in what is done, and the witness, like the witness at a wedding, takes no part but sees it and attests.

Heart, mind and consciousness (author's mapping, Interpretive). In plain words the three perspectives are the heart (I1), the mind (I2) and the eye of consciousness (I3). In the empathy triad (Table 2 gives the mode of each claim), emotional empathy is felt in the heart and body, cognitive empathy is considered in the mind, and compassion is the heart directed by the mind and acted on, and seen by consciousness when it is witnessed. These are perspectives, not parts of a person. The wish to help belongs to the felt side and understanding directs it; the soul is the name for the whole person acting on both, with the witness looking. "Spiritual" (I1, the felt inner life) and "soul" (that life joined with understanding, witnessed and acted on) are therefore different things here.

Identity and difference (Interpretive). I1 and I2 are perspectives whose settings can be measured, which the hypotheses below do; I3 is the eye they are seen through, so it is tried

rather than measured; the witness is the self looking through it. Read as the companion essay reads them (paper I; the Unifying Paper, part 2, A3), the pair is identity and difference: concern is the other known as not other, understanding is the other known as other, and compassion is identity directed by difference, the other known as oneself and as other in one act. Concern without understanding slides into contagion and, when it overwhelms, into distress, the other not other at all; understanding without concern into indifference or cruelty, the other only other; and the humanity paradox rests, in these terms, on the gap between knowing the other's mind and knowing the other as oneself, Schopenhauer's compassion through the seeing-through of separateness being the nearest older statement (Schopenhauer, 1840/1995). The ego, identity held in place by difference, forms in childhood, with self-recognition in the mirror from about eighteen months and the first self-conscious emotions soon after (Measured; Amsterdam, 1972; Lewis, 1992); on this reading compassion and cruelty both need that edge, since both need the other as other, and the way offered here is not back to the innocence before it but forward to I3, where the edge is seen and no longer load-bearing; a second innocence, with the edge seen, Ricoeur's second naïveté (Ricoeur, 1960/1967) (Interpretive).

Table 2. The three empathy claims, their mode and basis.

Claim	Mode	Basis
Emotional empathy is felt in the heart and body	Interpretive, with Measured support	Empathy for pain activates regions that track the body's own state, such as the anterior insula (Singer et al., 2004, 16 couples): the activation is Measured, and "in the heart and body" is a reading of it.
Cognitive empathy is considered in the mind	Interpretive, with Measured support	The two empathies dissociate in lesion work (Shamay-Tsoory et al., 2009) and cognitive empathy engages mentalizing networks (Zaki & Ochsner, 2012): the dissociation is Measured, and "in the mind" is a reading of it.
Compassion is the heart directed by the mind and acted on	Interpretive	A meaning given to the measured pieces; H1 tests the interaction of the two empathies, not the meaning.

This essay ties I1 to emotional empathy (feeling with another person) and I2 to cognitive empathy (working out what another person thinks and feels). **Compassion is proposed here as the seam where the two meet, the overlap that Figure 2 draws; I3 is the eye from which the seam is seen. This is an interpretive claim built on a measured distinction.** Singer and Klimecki's (2014) distinction (section 2) is the measured ground: feeling another person's pain without enough distance can turn into personal distress and withdrawal, whereas concern tends to sustain care. As a further reading (Interpretive), the I1 range is also how people attune: feeling spreads between people and matches (the matching half of empathy; Measured as emotional contagion, Hatfield et al., 1994), which on this reading is what draws like states together, while compassion answers unlike states too.

- **Emotional empathy alone (I1):** the felt connection is there, but it can overwhelm, and it can be narrow, favouring those who are close or similar.
- **Cognitive empathy alone (I2):** the understanding is there, but without felt concern it need not lead to care. I2 is neutral, so this is incompleteness, not a fault in I2. Blair (2005) reported intact cognitive empathy (theory of mind) alongside reduced emotional empathy in psychopathy (Measured). A 2025 meta-analysis qualifies that picture (Measured): psychopathic traits show the clearest link with reduced emotional empathy (r about -.35; Shukla & Upadhyay, 2025); the link with cognitive empathy is weaker (r about -.23) and varies widely across studies (I-squared about 94%), so "intact cognitive empathy" is too strong a claim. Here "relatively preserved" means cognitive empathy that is higher than emotional empathy in the same group, not normal or complete.
- **Compassion (I1+ directed by I2):** felt concern, directed by understanding, and acted on to help; witnessed, with I3 taken up, it can be noticed and chosen. The wish to help comes from the felt side and is what Singer and Klimecki call compassion (warm concern

with a wish to help), and concern here; understanding directs it. That is why compassion here is acted, and not only seen. Where the will to act ultimately comes from is not something this essay can measure; that part is an Experienced and Interpretive claim.

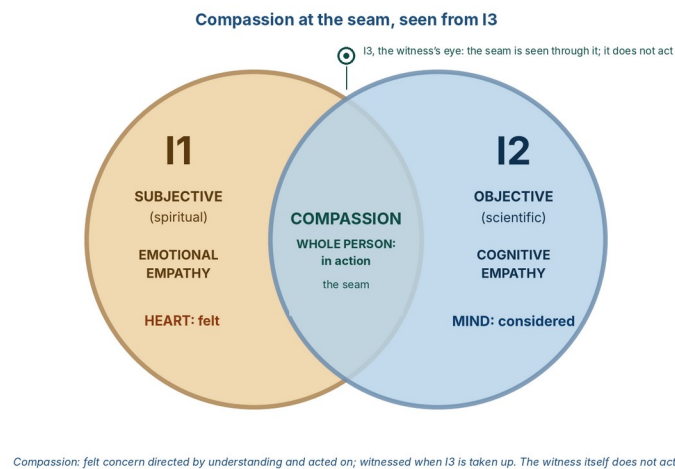


Figure 2. The two empathies and their seam (Interpretive). I1, the felt perspective (subjective), holds emotional empathy (heart); I2, the understanding perspective (objective), holds cognitive empathy (mind); Compassion, felt concern directed by understanding and acted on, is their overlap, the seam. I3, the witness's eye (transjective), is drawn above it as the eye from which the seam is seen; above on the page only, since the eye is not a higher level of reality, and it does not act. The overlap is drawn for the positive pole of I1; with I1- the same region holds cruelty and revenge, and defended detachment beside it (Table 3).

One limit: compassion is not a level above empathy. It is felt concern, directed by understanding, and acted on to help, which keeps it at the seam shown in Figure 2.

Six states, filling five of the nine combinations of a setting of I1 and a role of I2, are named in Table 3 (Interpretive; these are states a person can be in, not types of people). Each can be witnessed (I3 taken up, the witness looking through it) or not; the table names the states without the witness, and witnessing is what lets a person notice which row they are in and move. Of the other combinations, I1+ alone is love at rest, calm with concern present and nothing asked of it, equanimity, which is I1+ at rest and not neutral; neutral is calm with concern absent, and at its dead end dullness, the ingredient of indifference. I1+ with I2 and neutral I1 directed by I2 are possible but are not named here, nor are the appetitive form of I1- alone (impulsive appetite) and with I2, which share their cells with distress and defended detachment; neutral I1 alone is simply calm:

Table 3. Compassion, indifference, defended detachment, cruelty, revenge and distress, told apart by the setting of I1 and the role of I2 (Interpretive).

State	Combination	What it looks like
Compassion (+)	I1+ directed by I2	Felt concern, understanding, and action to help.
Indifference (0)	Neutral I1 with I2	Understanding is present but nothing is stirred and nothing is defended: the person is simply not moved.
Defended detachment (-)	I1- (reactive form) with I2	Understanding is present but follows the feeling, which is shut down or walled off for self-protection: numbing, avoidance.
Cruelty or exploitation (-, active)	I1- (appetitive form) directed by I2	Understanding sets the course and serves appetite or aggression, with felt concern for others absent: understanding put to work for calculated harm.
Revenge (-, injured)	I1- (reactive form) directed by I2	Hurt made into a plan: understanding sets the course for the wound, a demand for identity forced through difference ("now you know what I felt"); the cycle rebuilds, since the other's suffering is never one's own. Its opposite, forgiveness, is compassion's combination met with injury.
Distress (-, overwhelmed)	I1- (reactive form) alone	Overwhelmed by the other person's feeling or one's own, with I2 offline; tends toward withdrawal.

At the extreme, and when it is a stable pattern, the cruelty row overlaps with what clinicians call psychopathic or sociopathic traits, where the empathy research reports reduced emotional empathy with cognitive empathy relatively preserved (above). In this scheme that is the appetitive form of I1–: low on the fear and distress measures that mark the reactive form, and separated from neutral I1 by the presence of a self-serving drive, which is why Table 5 gives the two forms different indicators. The rows therefore differ by the form of I1– as well as by the role of I2, and the two forms are not one pole seen twice: distress and appetite can be measured apart. Table 3 names states, not diagnoses, and must never be used to label an individual.

Aggression research separates reactive aggression, which is hot and impulsive and answers threat or frustration, from proactive aggression, which is planned and instrumental (Measured; Dodge, 1991). In this scheme reactive aggression is the reactive form of I1– turned outward: impulsive when it governs alone, close to the distress row, and revenge when understanding directs it, and proactive aggression is the appetitive form of I1– directed by I2, the cruelty row. Proactive aggression has been linked to psychopathic traits and reactive aggression to impulsivity and hostility in adolescent boys (Measured; Raine et al., 2006). These too are states, not types, and one person can show both.

Used to describe states, the scheme earns its keep: most people move between them (Interpretive; H3 tests the nearest measurable version, whether a person's emphasis on feeling or on thinking behaves as a state or as a trait), and they differ in how easily they step back from their own reactions, which is what the witness provides: flexibility; the stance is tried, and its correlate, psychological flexibility, is measured in acceptance and commitment therapy as the inverse of experiential avoidance (Measured; Hayes et al., 2006). Until I3 is taken up, the person is simply in the state. In a mind racing with worry, the fear is I1– and the thoughts that go with it are I2, which may know better without steering: the reactive form of I1– with I2 but not directed, which Table 3 calls defended detachment; worry is that combination in a racing form, numbing and avoidance its usual one, since the scheme carries no measure of arousal. Used to sort people into kinds, it fails, since people move over a lifetime.

7. Testable hypotheses (proposed)

Most of this essay is Interpretive. Four parts of it can be tested. Table 4 sets out each one with its prediction, a design and what would count against it.

Table 4. The four hypotheses, with predictions, designs and what would count against them.

Hypothesis	Prediction	Design	Counts against
H1. Compassion sits at the seam of the two empathies	Helping or compassionate action is predicted by the interaction of emotional and cognitive empathy, beyond each alone.	Cross-sectional or experimental study in adults, with a validated scale that separates empathic concern, personal distress and cognitive empathy (for example the Interpersonal Reactivity Index, IRI; Davis, 1983) and a behavioural helping measure; one pre-registered moderated-mediation model, in which the indirect path (perspective taking to concern to helping) and the interaction of concern with cognitive empathy are estimated together, with personal distress as a covariate.	Only additive effects; one empathy alone accounts for the effect; or, with the indirect path in the model, the interaction's confidence interval includes zero or it adds less than an increment in explained variance fixed in advance.

Hypothesis	Prediction	Design	Counts against
H2. Labelling transfers	Readers given the definitions of the three modes and a passage whose statements are labelled Experienced, Measured or Interpretive classify a new set of unlabelled statements as report, finding or reading more accurately than readers given the same definitions and the passage unlabelled. The test items carry no labels, so the labels can help only if the distinction has been learned.	Pre-registered online experiment, between subjects: both groups first read Table 1's definitions, then the same passage, one version labelled and one not, then classify a fresh set of unlabelled statements, scored against classifications agreed in advance by independent raters. H2 therefore tests whether labelled examples add to the definitions.	No difference on the unlabelled test items, or labelled readers do worse.
H3. Within-person variation in emphasis	Reported emphasis on the felt perspective versus the understanding perspective varies more within a person across days than between people, over several weeks; the reading that people move between states over a lifetime is wider than this test.	Experience sampling over several weeks, with a short rating at each prompt of how far feeling or thinking led the person's recent thoughts and actions (a short set of items written for the study and piloted first for reliability, since the error of a single item would count as within-person variance); intraclass correlation, which divides the variation into within-person and between-person parts and shows how stable the emphasis is, not whether people fall into kinds.	Intraclass correlation above 0.5, meaning the emphasis behaves like a stable trait.

Hypothesis	Prediction	Design	Counts against
H4. Indifference and cruelty are different profiles	The patterns behind indifference and cruelty separate on behaviour, tested first as continuous moderation (concern by distress predicting helping; callous traits by cognitive empathy predicting costly harm) and then as three profiles specified in advance: a neutral profile (low emotional empathy, low distress and avoidance, and no self-serving drive) helps less than participants high in concern and low in distress, and harms no more than they do; a defended profile (high distress and avoidance, cognitive empathy relatively preserved; the reactive form of I1-) helps less than the same comparison group; a harm profile (callous traits on the Inventory of Callous-Unemotional Traits, with low distress, cognitive empathy relatively preserved and proactive aggression high; the appetitive form of I1-) takes costly harming actions more often. Proactive aggression on the RPQ defines the self-serving drive that separates the harm profile from the neutral one, so that the two are not two degrees of one empathy scale; it is not part of the outcome. Personal distress and avoidance are expected to track the defended profile; reactive aggression is expected to track distress and provocation rather than callous traits, so harm that follows provocation, revenge, is read apart from the harm profile, which is why the harm task records whether the participant was provoked.	Cross-sectional study with a validated scale that separates empathic concern, personal distress and cognitive empathy (the IRI; Davis, 1983), the Inventory of Callous-Unemotional Traits (Frick, 2004), a measure of experiential avoidance, a measure of reactive and proactive aggression (the RPQ; Raine et al., 2006), and two behavioural tasks, one for helping (the Redistribution Game; Weng et al., 2013) and one for costly harm (the Taylor Aggression Paradigm, with provocation recorded; Taylor, 1967); moderated regression as the primary test, and latent profile analysis as a secondary one, with decision rules fixed in advance (information criteria, a bootstrap likelihood-ratio test, smallest profile at least five per cent), in a sample enriched for callous traits by screening, since the harm profile is rare in community samples, with the estimates weighted back to the screened population so that the screening does not bias the moderation estimate.	The moderation terms do not predict the two tasks; or the profiles do not separate on them; or one empathy score, or the contrast of empathic concern with personal distress alone, predicts both helping and harm as well as the profiles do, by a difference in explained variance fixed in advance.

Planning sizes, from standard power calculations at 80 per cent power and a 5 per cent two-sided error rate: H1 needs about 400 adults to detect a small interaction (f -squared of 0.02), and in practice several times that, since interaction terms built from self-report scales are weakened by measurement error; H2 needs about 350 readers in total to detect a modest difference (d of 0.3) on the transfer items. The H3 cut-point (an intraclass correlation of 0.5) is conventional, not a result. H4's primary test is a moderated regression on continuous scores; the scores are dispositional, so H4 tests the dispositional correlates of the states, and H3 and H4 can both hold, since emphasis can vary within a person while the range a person can reach differs between people; its secondary latent profile analysis needs several hundred participants, a sample enriched for callous traits, and a power analysis specific to the design; no figure is given here. None of these studies has been run; Table 5 lists example instruments.

What rival accounts predict for H1. An empathic-concern account (Batson, 2011) predicts that emotional empathy alone predicts helping; a cognitive account, in the spirit of Bloom's (2016) case for reasoned compassion over felt empathy, predicts that cognitive empathy alone does; an additive account predicts two main effects and no interaction; and a mediational account (Coke et al., 1978; Davis, 1983), in which perspective taking arouses empathic concern and concern drives helping, predicts that the effect of cognitive empathy runs through concern, with no interaction. The design estimates the mediational path and the interaction in one model; only a positive interaction that survives with the path in place favours the claim, so H1 can fail in four ways, each informative.

Rival accounts for H4. Batson's empathy-altruism work already predicts the defended profile's result: in his easy-escape experiments, people who felt little empathic concern helped much less when escape was easy, whereas people who felt high concern helped either way, and people reporting distress rather than concern helped less when escape was easy (Batson, 2011; Batson et al., 1983). That part is not new; H4 adds the neutral and harm profiles. H4 does not manipulate escape, so its prediction for the defended profile is the simpler one, less helping than the comparison group, and Batson's result is the reason to expect it. The rivals are a one-dimension account, in which a single empathy score, or the contrast of empathic concern with personal distress, predicts both helping and harm; H4 fails if that account predicts helping and harm as well as the moderation terms and the profiles do, by a criterion fixed in advance. H4 does not test "overwhelmed" (I2 offline) against "defended" (I2 present); that needs a task that loads understanding, and is left for later.

Table 5. How the settings of I1, including the two forms of I1-, I2 and three of the six named states (compassion, cruelty and revenge) could be measured (proposed). The mapping from settings, perspectives and states to instruments is Interpretive; the instruments themselves are published measures.

Setting, perspective or state	What to measure	Example instrument or task
I1+ (felt concern)	Affective empathy and concern for others; felt connection	Empathic Concern subscale of the IRI (Davis, 1983); self-report compassion scales. The affective subscales of the QCAE (Reniers et al., 2011) do not separate concern from distress
Neutral I1	Low emotional empathy together with low distress and low avoidance	The same scales, with scores low on both; an experiential avoidance measure
I1- (reactive form)	Personal distress, avoidance, threat-driven reactivity; felt concern for others absent	Personal Distress subscale of the IRI (Davis, 1983); avoidance scales; reactive aggression on the RPQ (Raine et al., 2006)
I1- (appetitive form)	Callous, self-seeking drive with felt concern for others absent; low fear and low distress	The Inventory of Callous-Unemotional Traits (Frick, 2004); proactive aggression on the RPQ (Raine et al., 2006); low scores on the Personal Distress subscale of the IRI
I2 (understanding)	Cognitive empathy, working out what another person thinks and feels	Perspective Taking subscale of the IRI; cognitive subscales of the QCAE; mentalizing tasks
Compassion (I1+ directed by I2)	Helping that costs the helper and is aimed at the other person's need	Behavioural helping tasks such as the Redistribution Game (Weng et al., 2013)
Cruelty or exploitation (I1-, appetitive form, directed by I2)	Callous traits and planned, instrumental harm with cognitive empathy relatively preserved	Proactive aggression on the RPQ; the Inventory of Callous-Unemotional Traits; the Taylor Aggression Paradigm (Taylor, 1967)
Revenge (I1-, reactive form, directed by I2)	Harm planned in answer to injury, with the motive measured as the wish to see the offender hurt	Revenge subscale of the Transgression-Related Interpersonal Motivations inventory (TRIM; McCullough et al., 1998); reactive aggression on the RPQ; the Taylor Aggression Paradigm with provocation recorded

8. Objections and replies

- "The source is introspection and informal observation", "this cannot be tested" and "mixing science and spirit is a category error." These three objections are answered in full in the companion essay (paper I, section 7). Briefly: the position statement says what the sources are and why they are limited, which is why the testable parts are stated in section 7; the Interpretive claims are labelled and judged on coherence and usefulness; and the labels exist so that a reader always knows which kind of claim is being made, so neither language is asked to do the other's job.

- **“Evolutionary mismatch is speculative.”** The essay includes Zuk’s caution and labels mismatch claims accordingly.
- **“The I1, I2 and I3 scheme is only a relabelling.” Largely so: it is an organising scheme, and it is Interpretive. Its testable content is H1 (compassion at the seam), H3 (within-person variation) and H4 (indifference and cruelty as different profiles); if all three fail, the scheme should be kept, if at all, as a metaphor and not as an account of how compassion arises.**

Risks of the framework

The framework could be misread in four ways: as presenting spiritual claims as findings; as encouraging people to skip medical or psychological care, which nothing here does; as the view of one person with a stake in it; or as a claim that some people are born for harm, which the humanity paradox does not make: it is about a fraction of strategies in a population, never about an individual.

9. Discussion

Read through the humanity paradox, the essay’s answer has two parts, and it is Interpretive. In the person, compassion is the state in which the two empathies are rejoined, felt concern directed by understanding, so that understanding serves others. In the population, the few who put understanding in the service of the negative pole are contained not by felt concern, which extended without understanding is what they exploit, but by the structures of section 5: repeated dealing, visible reputation, monitoring and fair rules, understanding made durable (Ostrom, 1990). Compassion without those structures is exploitable; structures without compassion drift toward the indifference row. The claim is that both are needed, and neither is derived from the other. Paper VI (section 7) carries the two halves to machines and to all conscious beings.

Implications. If concern, in Singer and Klimecki’s sense, is a trainable response that differs from empathy (Klimecki et al., 2014; Weng et al., 2013), programmes in education, healthcare and clinical settings might do better to treat the two separately: building understanding and a way of staying with another person’s feeling without being overwhelmed, not only raising empathic arousal. Where a person is mostly in the distress state of Table 3, the aim would be to restore enough distance for understanding to engage; where understanding is present but nothing is stirred, the aim would be to cultivate felt concern. The labelling discipline could also help interdisciplinary teaching and science communication, where findings and metaphors are easily confused. These are suggestions for study, not recommendations for practice, and nothing here replaces clinical care.

Limits. No new data are reported, and the scheme rests on one author’s introspection and informal observation. Most of the essay is Interpretive, and the scheme of perspectives (I1, I2, I3) has measurable content only through H1, H3 and H4. The tests of H1 and H4 reach only the empathy pair (emotional and cognitive empathy); the wider content of I1, such as conscience, grace and felt connection, stays Experienced and Interpretive. Most of the evidence cited is from Western samples, and the scheme itself was drawn in one, so neither may generalise. H1 tests the first half of the essay’s answer to the paradox, compassion in the person; the second half, the structures of section 5, is not tested here, and H4 tests the profile distinction the population claim relies on; none of them tests the paradox itself, which rests on the Measured dissociation of the two empathies and on the modelled frequency dependence of section 5. A stable trait component under H3 would not refute the paradox.

What would change the view. A failure of H1 would remove the claim that compassion sits at the seam of the two empathies. A failure of H2 would show that the labels do not transfer, leaving the discipline a convention rather than a teachable skill. A failure of H3 would show that a person’s emphasis is stable over weeks, trait-like rather than state-like, which would narrow the claim that these are states a person moves between. A failure of H4 would count against the claim that the usual profiles corresponding to indifference and cruelty differ in what people do.

Next steps. The cheapest test is H2 (does labelling transfer to statements that carry no label), which needs about 350 participants in an online experiment; paper VI (section 8) runs the same test on a machine's outputs as its H1. H1 can be run on existing empathy and helping measures. Anyone is welcome to test, extend or refute any part.

One way to close. Be careful not to identify with the state you are in, for the centre, the source of your own light, is but a moment of experience away from being realised. That is advice, and sits outside the modes; the states are those of Table 3, and the way offered to the centre is the companion essay's (paper I, section 6). Measurement tells us how far a claim can be checked; some of what matters here is met only by living it, and the essay has tried to keep both in view without letting either pretend to be the other.

Your centre, our centre, the centre of all.

Declarations

Competing interests. The author founded Centricity, a community project about neighbours sharing tools, skills and help. It is a separate project: it pays the author no salary or other income; the systems paragraph in section 5 states the thinking behind it, but the project is not part of this essay's hypotheses and the essay does not depend on it.

Funding. No funding, grant or employer support was received for this work.

Author contributions. The author conceived the framework, did the research and wrote the manuscript.

Ethics. No new data were collected from human participants. No ethics approval was needed.

Data availability. No new data were generated or analysed. All numbers cited are from the published sources in the reference list; Figures 1 and 2 are the author's own diagrams.

Use of AI. Generative AI (Claude, from Anthropic: Sonnet 5.5 for drafting, Fable 5.1 and Opus 5.5 for revision; Anthropic, 2026) was used to help draft and edit the text, to render the figures from the author's own sketches and descriptions (both are original diagrams; neither is reproduced from another publication), and to check references and internal consistency. The author directed the work, reviewed all content, verified the cited sources and takes full responsibility for it. No AI tool is an author.

References

- Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology*, 5(4), 297–305.
<https://doi.org/10.1002/dev.420050403>
- Anthropic. (2026). *Claude* (Sonnet 5.5, Opus 5.5 and Fable 5.1 versions) [Large language model].
<https://claude.ai>
- Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.
- Batson, C. D. (2011). *Altruism in humans*. Oxford University Press.
- Batson, C. D., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of Personality and Social Psychology*, 45(3), 706–718. <https://doi.org/10.1037/0022-3514.45.3.706>
- Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition*, 14(4), 698–718.
<https://doi.org/10.1016/j.concog.2005.06.004>
- Bloom, P. (2016). *Against empathy: The case for rational compassion*. Ecco.
- Byrne, R. W., & Whiten, A. (Eds.). (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans*. Clarendon Press.

- Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiandreou, E., & Inzlicht, M. (2019). Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General*, 148(6), 962–976. <https://doi.org/10.1037/xge0000595>
- Coke, J. S., Batson, C. D., & McDavis, K. (1978). Empathic mediation of helping: A two-stage model. *Journal of Personality and Social Psychology*, 36(7), 752–766. <https://doi.org/10.1037/0022-3514.36.7.752>
- Crusio, W. E. (1995). The sociopathy of sociobiology. *Behavioral and Brain Sciences*, 18(3), 552.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126. <https://doi.org/10.1037/0022-3514.44.1.113>
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2), 71–100. <https://doi.org/10.1177/1534582304267187>
- Dodge, K. A. (1991). The structure and function of reactive and proactive aggression. In D. J. Pepler & K. H. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 201–218). Lawrence Erlbaum.
- Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. W. W. Norton.
- Frick, P. J. (2004). *Inventory of Callous–Unemotional Traits* [Unpublished rating scale]. University of New Orleans.
- Gluckman, P., Beedle, A., & Hanson, M. (2009). *Principles of evolutionary medicine*. Oxford University Press.
- Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin*, 136(3), 351–374. <https://doi.org/10.1037/a0018807>
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859), 1243–1248. <https://doi.org/10.1126/science.162.3859.1243>
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge University Press.
- Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, 44(1), 1–25. <https://doi.org/10.1016/j.brat.2005.06.006>
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of Homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2), 73–78. <https://doi.org/10.1257/aer.91.2.73>
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327(5972), 1480–1484. <https://doi.org/10.1126/science.1182238>
- Heym, N., Kibowski, F., Bloxsom, C. A. J., Blanchard, A., Harper, A., Wallace, L., Firth, J., & Sumich, A. (2021). The Dark Empath: Characterising dark traits in the presence of empathy. *Personality and Individual Differences*, 169, Article 110172. <https://doi.org/10.1016/j.paid.2020.110172>
- Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 303–317). Cambridge University Press.
- Imhof, L. A., Fudenberg, D., & Nowak, M. A. (2005). Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences*, 102(31), 10797–10800. <https://doi.org/10.1073/pnas.0502589102>
- Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. *Social Cognitive and Affective Neuroscience*, 9(6), 873–879. <https://doi.org/10.1093/scan/nst060>
- Lewis, M. (1992). *Shame: The exposed self*. Free Press.
- Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press.
- Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, 246(5427), 15–18. <https://doi.org/10.1038/246015a0>
- McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Jr., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, 75(6), 1586–1603. <https://doi.org/10.1037/0022-3514.75.6.1586>
- Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences*, 18(3), 523–541. <https://doi.org/10.1017/S0140525X00039595>
- Nesse, R. M., & Williams, G. C. (1994). *Why we get sick: The new science of Darwinian medicine*. Times Books.

- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805), 1560–1563. <https://doi.org/10.1126/science.1133755>
- Nowak, M. A., & Sigmund, K. (1992). Tit for tat in heterogeneous populations. *Nature*, 355(6357), 250–253. <https://doi.org/10.1038/355250a0>
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1–20. <https://doi.org/10.1017/S0140525X02000018>
- Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., Stouthamer-Loeber, M., & Liu, J. (2006). The Reactive–Proactive Aggression Questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior*, 32(2), 159–171. <https://doi.org/10.1002/ab.20115>
- Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A Questionnaire of Cognitive and Affective Empathy. *Journal of Personality Assessment*, 93(1), 84–95. <https://doi.org/10.1080/00223891.2010.528484>
- Ricoeur, P. (1967). *The symbolism of evil* (E. Buchanan, Trans.). Harper & Row. (Original work published 1960)
- Schopenhauer, A. (1995). *On the basis of morality* (E. F. J. Payne, Trans.). Berghahn. (Original work published 1840)
- Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, 132(3), 617–627. <https://doi.org/10.1093/brain/awn279>
- Shukla, M., & Upadhyay, N. (2025). Cold hearts and dark minds: A systematic review and meta-analysis of empathy across dark triad personalities. *Frontiers in Psychiatry*, 16, Article 1546917. <https://doi.org/10.3389/fpsyt.2025.1546917>
- Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology*, 24(18), R875–R878. <https://doi.org/10.1016/j.cub.2014.06.054>
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661), 1157–1162. <https://doi.org/10.1126/science.1093535>
- Sosis, R., & Bressler, E. R. (2003). Cooperation and commune longevity: A test of the costly signaling theory of religion. *Cross-Cultural Research*, 37(2), 211–239. <https://doi.org/10.1177/1069397103037002003>
- Summerhayes, D. (2026). *Arriving at the centre of consciousness: A transjective framework* [Manuscript submitted for publication].
- Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality*, 35, 297–310. <https://doi.org/10.1111/j.1467-6494.1967.tb01430.x>
- Vervaeke, J. (2019, August 16). *Ep. 31 – Awakening from the meaning crisis – Embodied-embedded RR as dynamical-developmental GI* [Video]. YouTube. <https://www.youtube.com/watch?v=gfKcVbNd7Xc>
- Viding, E., Blair, R. J. R., Moffitt, T. E., & Plomin, R. (2005). Evidence for substantial genetic risk for psychopathy in 7-year-olds. *Journal of Child Psychology and Psychiatry*, 46(6), 592–597. <https://doi.org/10.1111/j.1469-7610.2004.00393.x>
- Weng, H. Y., Fox, A. S., Shackman, A. J., Stodola, D. E., Caldwell, J. Z. K., Olson, M. C., Rogers, G. M., & Davidson, R. J. (2013). Compassion training alters altruism and neural responses to suffering. *Psychological Science*, 24(7), 1171–1180. <https://doi.org/10.1177/0956797612469537>
- Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience*, 15(5), 675–680. <https://doi.org/10.1038/nn.3085>
- Zuk, M. (2013). *Paleofantasy: What evolution really tells us about sex, diet, and how we live*. W. W. Norton.