The ETHIC Stack: The Neglected Mechanisms of Moral Decision-Making (WIP)

A multi-level, mechanistic framework for ethical competence in warfighting

Dorian Minors

Abstract

Ethical decision-making has been, and remains, a critical feature of modern warfighting. This is not simply because it is a humanistic imperative, but because it is a combat-power multiplier. Yet two chronic defects plague most doctrinal and Professional Military Education (PME) approaches to the subject. The first is an overemphasis on "reason"—deliberative, or evaluative processes of decision-making. The second is that they typically remain prescriptive rather than mechanistic. To address these issues, I sketch a tentative five-layer mechanism of moral action that lays the causal plumbing required to intervene on and improve ethical judgement. Specifically, I suggest what happens at the neural ([E]arly emotional), cognitive ([T]hought-level schemas), situational ([H]abitat), relational ([I]n-group dynamics), and cultural ([C]ultural and institutional scaffolding) levels, that inform moral behaviour. Hence, the ETHIC Stack, a multi-level, mechanistic framework for ethical competence in warfighting.

Contents

1 Introduction

A note of special thanks to Major Benjamin Ordiway, whose Moral Terrain project inspired this article, and several of its features.

<Vignette about Montgomery, Intake 1 at Sandhurst on ethical leadership>.

Ethical decision-making has been, and remains, a critical feature of modern warfighting. This is not simply because it is a humanistic imperative, but because it is a combat-power multiplier. Specifically, judgements involving moral content made "in the moment" do not just decide tactical success, they also determine post-conflict legitimacy. This is why the "moral component" of fighting power is judged "arguably the most important" in British leadership doctrine, because "success on operations is dependent upon people to a greater degree than equipment or tactics" (Centre for Army Leadership (CAL), 2021, p. 11). This perspective is not restricted to the British, of course. Ethical leadership "is the single most important factor in ensuring the legitimacy of [Australian] operations and the support of the Australian people" in Australian leadership doctrine (Lessons and Doctrine Directorate ADF, 2021, p. 7) for example, and reflections from our U.S. partners value "the [U.S.] Army Ethic" as their "greatest asset, even when compared to the technology and weaponry of the modern operating environment", critical to both the morale of soldiers and trust of the U.S. populace (Crayne, 2025, p. 6). It is no surprise, therefore, that models of ethical decision-making proliferate among military establishments.

Sadly, two understandable, but chronic defects plague most doctrinal and Professional Military Education (PME) approaches on the subject that make their utility questionable.

The first is an overemphasis on "reason"—deliberative, or evaluative processes of decision-making. As Section 2 of this article explores in more detail, deliberative thought appears to more frequently rationalise established intuitions, desires, and beliefs, as opposed to challenge or adjust them. Troublingly, moral judgements seem particularly vulnerable to

this property of human cognition. More troubling still, even when this is not the case, deliberative processes of thought are particularly at risk when stress, time-pressure, fatigue, or habituated brutality predominate—precisely those moments where deliberative ethical competence might be most valuable.

The second defect in many efforts to improve moral decision-making in the military sphere is that they typically remain prescriptive rather than mechanistic. As Section 3 details more thoroughly, telling people what they ought to do has little utility unless one also demonstrates how they might achieve it. Perhaps more importantly, the "mechanistic turn" in cognitive science has been underway for at least a quarter-century—launched ostensibly by Machamer, Darden, and Craver's (2000) seminal work and elaborated in, for example, Craver's (2007) account of neural mechanisms, Bechtel's (2007) analyses of mental mechanisms, and Glennan's (2017) multi-level mechanisms—yet research on military ethics has yet to absorb these insights in earnest.

Two illustrative cases show how the defects surface in practice: Kem's "Ethical Triangle" (2006), dominant in the U.S., and Sandhurst's own S-CALM model (Vincent, 2022). Each are elegant blueprints for ethical decision-making in the military context that are, nevertheless, limited thus.

I will go on, in Section 4 of this article, to sketch in rough the kind of framework I suspect would remedy these defects. Grounded in concepts drawn from the mechanistic turn in cognitive and social sciences, I will propose a multi-level approach to ethical competence that both accounts for the more inconvenient properties of human decision-making and lays the causal plumbing required to intervene on and improve ethical judgement.

To conclude, I will illustrate how: the framework neatly accommodates and augments existing models of ethical decision-making; supports and sharpens recent innovations in PME recommendations around ethical education; and lends itself to—and in fact, necessitates—the adaptation of these things for cultural variation in norms and values; before highlighting those areas where this rough framework is weak and requires improvement. In particular, I desire to make clear that, in the language of the "New Mechanists", this sketch is a "how-possibly" explanation of moral behaviour, not a "how-actually" one,

laying groundwork for development into something less possible, and more plausible.

2 Deliberation may be more slave than master

There is a tension in the way in which we grapple with the role of deliberative thought in shaping human behaviour. On the one hand, deliberative thought is frequently interpreted to be the necessary master of our more emotional intuitions. Epictetus paired emotionally coloured impressions (*phantasia*) with our power to rationally decide whether those impressions were true and thus to act on them (*prohairesis*) in *Discourses*. Aquinas paired the *appetitus sensitivus*, the sense appetite, with the *appetitus rationalis*, the rational appetite in *Summa Theologiae*. Kant laid out his Categorical Imperative in *Groundwork for the Metaphysics of Morals*. Kahneman posed his 'fast' cognitive System 1 against his 'slow' cognitive System 2 in *Thinking Fast and Slow*.

On the other hand, there is the uncomfortable understanding that, at least sometimes, deliberative thought is a slave to these very same emotional intuitions. Aristotle noted in *Nicomachean Ethics* that when appetite wins over reason, we will counterfeit syllogisms to excuse the victory of passion. Hume, in the *Treatise of Human Nature*, pointed out that reason is the slave of the passions. Nietzche in *Geneology of Morality* illustrated that arguments are masks which our drives deploy after the fact. Damasio in *Descartes' Error* suggested that emotion provides the necessary signals which guide 'rational' choices.

2.1 We prefer deliberation over intuition

The stronger impulse appears to be to optimistically defer to the former—deliberation as master. Perhaps Mercier and Sperberger (2017) overstate the case, but they are certainly not wrong when they write in their introduction:

> Psychologists generally recognize that reason is biased and lazy, that it often fails to correct mistaken intuitions, and that it sometimes makes things worse. Yet most of them also maintain that the main function of reason is to enhance individual cognition

This is a major project of behavioural economics, for example (Kahneman, 2011; Thaler & Sunstein, 2009; though c.f. Maier et al., 2022), and remains a foundational premise of modern cognitive-behavioural therapeutic approaches (i.e. Beck, 1979; Ellis, 1962; see Hofmann et al., 2012). Mercier and Sperberger themselves centre their argument around the fact that reasoning in groups is quite reliable, in an effort to rescue the process.

Military ethicists similarly appear, on average, to favour deliberative processes of thought as a means of navigating moral terrain (though c.f. Messervey et al., 2021; Ordiway, 2022). For example, Kem's (2006) "Ethical Triangle Decision-Making model" has been described as "perhaps the most recognizable tool for ethical decision-making in the military" (Ordiway, 2022). In it, Kem suggests that the decision-maker's "courses of action should be tested against three completely different criteria for ethical decision making. They are: principles or the rules-based approach; consequences or the utilitarian approach; and virtues. These are the three basic schools of thought for ethics" (p. 27).[1] Nowhere does Kem engage with the fact that both his model, and the model upon which his work is based, will be subject to the inherent dangers in relying on such deliberative 'testing' in cases where deliberation is the slave of intuition.

We might also illustrate with an ethical decision-making model developed at the Royal Military Academy Sandhurst: Vincent's (2022) S-CALM model. As is perhaps appropriate for a model developed 16 years later than Kem's, Vincent more explicitly engages with the role of both deliberative and intuitive decision-making in moral judgement. The core premise of the model is that chains of unethical behaviour are dominated by intuitive, fast thought processes (p. 18). Vincent identifies thirteen of these, and highlights five situational factors that appear to influence them. He explicitly recommends that leaders should use their deliberative processes to "recognise how the [situational factors] may be driving their thought processes" (p. 32) and describes three lines of questioning to counteract unethical patterns that might ensue. Again advancing beyond Kem, Vincent goes as far as recognising that deliberative thought can also be error-prone. However, this

[1] Here, Kem is referring to deontological ethics, a combination of consequentialist and utilitarian ethics, and virtue ethics respectively, for those who wish to explore them further. Contra Kem's claim that these are the 'basic' schools of thought, there are in fact many more. One might also consider, for example, care ethics, communitarian and role ethics, moral particularism, and pragmatic ethics as a start.

insight only appears in a single aside in his seminal paper (p. 14).

Both models neatly illustrate the point that military ethics, like the sciences of mind, frequently rely on the promise of deliberative processes of thought to enhance decision-making. This seems entirely appropriate. Deliberative, conscious processing *does* adjust moral judgement (e.g. Cushman et al., 2006; Greene, 2014a; Musschenga, 2008) Unlike the sciences of mind, however, there is little explicit consideration of the fact that, while deliberation may sometimes serve as master, it also plays a role as slave.

2.2 Moral judgements are particularly vulnerable to intuition

Sadly, the weight of evidence, particularly when it comes to *moral* judgement, appears to favour the reverse—deliberative processes appear to more frequently rationalise our intuitions, rather than challenge or adjust them.

The obvious starting place to illustrate this would be Haidt's Social Intuitionist approach to moral judgement (Haidt, 2001; or Haidt, 2012 for the more casual reader). Couched in dual-process frameworks of cognition—theories which attempt to formally and empirically distinguish the kinds of deliberative and intuitive processes of thought we have been speaking of—-Haidt proposes:

> (a) [t]here are two cognitive processes at work – reasoning and intuition – and the reasoning process has been overemphasized, (b) reasoning is often motivated, (c) the reasoning process constructs *post hoc* justifications, yet we experience the illusion of objective reasoning; and (d) moral action covaries with moral emotion more than with moral reasoning (Haidt, 2001, p. 815).

Not everyone agrees with Haidt's damning interpretation of the moral failings of 'reason' (e.g. Brand, 2016; Greene, 2014b; Karssing, 2023; Musschenga, 2008), but his project has been taken very seriously by some military ethicists (see Karssing, 2023; Messervey et al., 2021; Ordiway, 2022 for good, compact overviews).

2.3 Moral deliberation is particularly compromised under stress

The concern of military ethicists is motivated in large part because deliberative processes of thought are particularly at risk of compromise when stress, time-pressure, fatigue, or habituated brutality predominate—precisely those moments where deliberative ethical competence might be most valuable (e.g. Messervey et al., 2023; Schwabe & Wolf, 2011; Shields et al., 2016; Starcke & Brand, 2012).

More problematically still, for those who favour deliberation over intuition, is the increasingly hard-to-ignore problem that reducing the complexity of human-decision-making to two categories leaves uncomfortable gaps. Models which find it necessary to distinguish deliberation and intuition from a third algorithmic or rule-following process aren't uncommon (e.g. Sauer, 2019; Stanovich, 2009). Other authors ignore the distinction entirely, making the case that reason is simply one of any number of processes involved in decision-making (e.g. Brand, 2016). Indeed, some single out the fact that, on this multi-process account, it makes more sense to interpret the role of deliberation as a tool developed to specifically process information in accordance with an initial belief, rather than challenging or adjusting those beliefs (e.g. Mercier & Sperber, 2017; see also Oeberst & Imhoff, 2023).

The upshot is that there is no particular reason to remain sanguine about the dominance of deliberation over more automatic and emotional intuitions. At best deliberation serves a dual role as master and slave. More likely, it is merely one of many competing cognitive processes. At worst, it is more slave than master.

3 Prescription fails under fire

A quarter-century ago, Machamer, Darden, and Craver (2000) launched a movement in philosophy of science when they made a seemingly common-sense observation: any model that hopes to account for reliable changes in the world must identify those things that actually produce the phenomenon of interest. Bechtel (2007), Craver (2007), and

Glennon (2017), among others, later extended the idea into the cognitive, neural, and social sciences: explanations that permit manipulation of the mind, brain, or people are those which explain *how* those things can be manipulated.

I would be surprised to find that the average reader disagreed with this conclusion, even without brushing up on mechanistic philosophy. Yet, it would not be controversial to say that the vast majority of military guides to ethical behaviour take the form of check-lists, flow-charts, or maxims. These documents declare what a soldier should do without spelling out how they might actually go about achieving the prescribed course of action. Indeed, in keeping with the deliberative optimism described in Section 2, much Professional Military Education (PME) still treats ethics as a collection of external rules to be consulted, rather than a system upon which one can intervene.

3.1 Prescription does not tell you *how*

This kind of prescription without mechanism—a causal chain linking stimulus to perception, appraisal, action, and feedback—runs the risk of disguising our inability to intervene on the causes of unethical behaviour—the *feeling* of understanding is under no obligation to reflect *actual* understanding (Davis, 1971; Trout, 2002; Wilson & Keil, 1998). In this particular case, normative prescriptions (rules) paired with procedural directives (check lists) *appear* to indicate a solution—you know what you should do, and what to do about it—but this process is under no obligation to bear fruit.

A brief tour of recent military scandals makes the problem concrete. Participants in the events at U.S. run Abu Ghraib prison, our own "Helmand Province Killing", or the Canadian "Somalia Affair"[2] were not ignorant of the Law of Armed Conflict; soldiers had been briefed, and laminated cards spelling out the rules were literally clipped to flak jackets. Yet, knowledge of rules did not translate into procedural action because the pre-

[2]For the unfamiliar reader, I will briefly describe these cases by matching in tone their Wikipedia entries for ease of internet search, though one could easily replace them with other examples. Abu Ghraib refers to a series of human rights violations and war crimes against detainees in the Abu Ghraib prison during the Iraq war. The "Helmand Province Killing" refers to the 2011 manslaughter of a wounded Taliban insurgent by Alexander Blackman. The "Somalia Affair" refers to the 1993 beating of Shidane Arone, a Somali teenager, to death by two Canadian peacekeepers.

scriptions never penetrated the causal chain that links occasion to perception, through thought, and onward into behaviour. In a 2014 conference held at Sandhurst, the "consensus was that while [British Army Values and Standards] in combination with [the Law of Armed Conflict] was absolutely essential, they are not sufficient in equipping officers and soldiers with the necessary 'tool box' with which to deal with 'real world' ethical problems in complex operational environments" Vincent (2022).

3.2 Prescription cannot resolve value conflicts

If we return to the two ethical decision-making models we used to illustrate Section 2, we see that they each emphasise the failings of prescriptive ethical logic. To start, recall Kem's "Ethical Triangle" (2006): a decision-maker should test a course of action against "three basic schools of thought for ethics"— (p. 27). For the first—rules-based ethics—Kem suggests one ask the questions "what rules exist" and "what are my moral obligations?" (p. 29). For the second—consequence-oriented ethics—Kem suggests one ask "what gives the biggest bang for the buck" and "who wins and loses?" For the final school—virtue-oriented ethics—Kem suggests one ask "what would my mom think?" or "what if my actions showed up on the front page of the newspaper?" (p. 32) Then, the individual should go on to select the course of action "that best represents Army values" (p. 33).

Initially, Kem's prescriptions seem, and are in many cases, quite helpful. It is an eminently memorable triage tool. However, we only have to look at Kem's own motivation to identify a difficulty in application. Kem explicitly highlights Kidder's (1995) observation that ethical values frequently come into conflict as an ambiguity the Triangle should resolve. Kidder noted that humans often face a choice between justice and mercy, for example, or between truth and loyalty. Kem suggests that we define ethical problems in such terms in order to test them "against ethical standards" (p. 26). Yet, the moment Kem's lenses disagree, as they predictably will, the model falls silent. If "what rules exist" conflicts with "what my mom might think" the Triangle provides no means to weight principles against virtues or consequences. It provides no cues for escaping the kinds of motivated reasoning described earlier. It recommends no training protocol for cultivat-

ing the requisite judgement under operational stress. We've simply reconstituted Kem's concern with different labels.

3.3 Prescriptive ingredients do not make a meal

Vincent's (2022) S-CALM framework makes great strides towards mechanism. Vincent takes pains to detail five situational factors and thirteen common, literature-derived behaviours that appear to be particularly associated with unethical behaviour. Yet, once these elements are laid out, and some levers highlighted, he reverts to prescription: leaders must "recognise" and "mitigate" these things, "apply their leadership skills", "use their moral compass" (p. 32) and measure decisions against the "useful check list" of the British Army Standards (p. 29). In other words, much of the causal plumbing remains unspecified. In particular, Vincent leaves unspecified the conditions under which situational factors interact with common behaviours to produce unethical behaviour, and thus when and how a leader might intervene. The capacity to "recognise cognitive changes" is, rather, treated as a given. More, "mitigation" is portrayed as an almost undifferentiated act of ethical leadership, but the eighteen (five plus thirteen) risk factors are hardly interchangeable. Vincent details their differences well, and highlights some levers for change, but his project does not reliably map risk factors to their corresponding corrective. Essentially, Vincent catalogues ingredients of ethical failure, but does not quite wire them together into a manipulable system.

Prescription without interconnection is not enough. A genuinely mechanistic account must specify not merely the ingredients—be they rules, virtues, situational factors, biases—but also the ways in which those ingredients interact to generate ethical or unethical action. In Section 4 I sketch such an account. Drawing on the "New Mechanist" idea that explanations hinge on entities (parts) and the activities that connect them, I propose a multi-level framework in the vein of Glennan (2017) that (1) identifies the causal stages in moral behaviour, (2) locates intervention points along that chain, and (3) offers practical levers for commanders who must make moral judgements in real time, or under complex circumstance.

4 The ETHIC Stack: a multi-level mechanistic framework

Now we must turn our attention from the easy thing—criticising other peoples' valiant attempts to tackle the problem of ethical decision-making—and do something to resolve the fissures we have exposed. Here I will attempt to describe a provisional causal story that wires together both the ingredients and those ways in which the ingredients interact to produce moral action. In the effort, I wish to be clear at the outset that my aim is simply to sketch the kind of thing I have in mind rather than attend to every philosophical and empirical nuance. Largely, this takes the form of me roughly formalising my claims with an illustrative formula, based on literature, to demonstrate that we must go further than simply handwaving, or risk the very prescription we are seeking to avoid. The detail matters, if we hope to help ethicists sharpen their own tools.

We should start by describing in detail what the New Mechanists would consider a mechanism (Machamer et al., 2000)—in my moral mechanism, I must attend to:

1. the mechanism's "entities": the (relatively) stable ingredients, or parts of the system;
2. it's "activities": the *things* that those parts engage in; and
3. the organisation: the way those things are linked together.

Usefully, one can illustrate these features using any kind of well understood contraption. So in the handheld radio sets any Officer should be familiar with, the antenna is an entity that captures the radiowave (activity). The tuner (entity) filters the wave to a specific frequency (activity). The heterodyning circuit (entity) drops that frequency into the audible range (activity). The amplifier (entity) boosts the signal (activity). The speaker (entity) converts it to sound (activity). The overarching organisation for this mechanism is clear: radiowaves are transformed into audible, intelligible sound.

The explanatory value of this organised schematic is that it shows precisely where one can "wiggle" something to change the outcome of the process. For example, as the sol-

dier who has had the reckless audacity to patrol into the hills has discovered, swapping her now useless short antenna for a longer one will help capture the lost radiowaves more effectively. Woodward (2005) calls this the "Manipulability Conception of Causal Explanation" (p. 9), and it is precisely the attribute missing from accounts of ethical behaviour I complained about in the previous section.

In what follows, I will outline a tentative five-layer mechanism of moral action that explicitly follows this parts-and-activities template, and how each layer cascades into the others. The levels are selected for when the dominant manipulable organisations shift from one domain to another—neural, cognitive, situational, relational, and cultural. In the game tradition of military acronymic zeal, I will call it *the ETHIC Stack*,[3] though this is more for fun than anything else and might actually make it less comprehensible:

- E: the early, emotional circuitry;
- T: the thought patterns and cognitive-social schemas;
- H: the immediate habitat (really, the situation);
- I: the in-group and social dynamics;
- C: the cultural, command, and institutional scaffolding.

At each level, I will describe in broad terms the entities, activities, and organisation that seem to characterise the level based on an exploratory account of relevant, candidate theories from the literature. I will then sketch a specific account of the mechanism at that level, drawing on those theories, along with a related intervention. In this way, though the model is exploratory—a "how-possibly", and sometimes "how-plausibly", but not yet "how-actually" blueprint—it is nevertheless mechanistic enough to guide both empirical research and practical training in ethical competence in the military context.

4.1 Early, emotional circuitry

We'll start where many, though not all, military ethicists do not—by describing the most basal, "intuitive" contributions to the mechanism. The candidate theories overviewed in Box 1 imply that the **entities** operating at this level would be the visceral and affective

[3]We must thank, or blame, AI for this. I can take no credit.

valuation systems of the nervous system. The **activities** they would engage in would be autonomic arousal and rapid affective tagging of stimuli that produce our intuitions. They likely also include the reinforcement learning loops that fix our habits and routines. The **organising principle**—the way in which the two are linked—would be the habits and hot-valuation processes which hand the higher-levels a valenced (positive or negative) state of arousal along with any associated action routines.

4.1.1 A mechanistic sketch

Reminder: I highlight this mechanism because it seems to be the most obvious intervention point, but other plausible mechanics are overviewed in Box 1.

Deriving from the literature overviewed in Box 1, and indeed any basic neuroscience textbook, we might sketch some basic features of a plausible mechanism. An abrupt interruption to your expectations—say, a door slamming—is going to noticed by your salience network within roughly a tenth of a second. Almost immediately, your affective network is going to tag the interruption with a valence—avoid (threat/disgust) or approach (care/protect). That will kick off an autonomic surge of arousal that reflects the scale of the interruption and is amplified by any stress already in the system. This all influences a dorsal-striatal queue of possible basic motor responses—whichever instinctive or otherwise over-learned responses are associated with the valence. The valence is passed on to the next level of the ETHIC stack, while the arousal determines which of the possible motor commands are issued—whatever can be "purchased" with the arousal.

We can spell the mechanism out a little more rigorously. Think of every over-learned or instinctive script as a candidate motor plan (`M_k`), each with its own "activation cost" `cost(M_k)`. The autonomic arousal level A first acts as the currency, "purchasing" scripts that have a cost matched by the energy currently in the system.

```
K_A = { M_k : cost(M_k) less than or equal to A }
```

From the "purchased set" of scripts (`K_A`), the system then chooses the specific plan that best fits the current valence `V`—positive biases towards approach-type scripts, or negative towards avoidance-type scripts:

```
M_chosen = max { g(A,k) * h(V,k) }
```

Where `g(A,k)` represents a 1 (the script `k` can be purchased with `A`), or a 0 (you can't afford the script)[4], and so it is either viable or not. This is multiplied by an emotional score `h(V,k)`, where the more the valence `V` (approach or avoid) matches the script `k`, the higher the score. The script with the highest score is the motor plan that's chosen (`M_chosen`).

Now, we're unlikely to be able to manipulate the valence very much at this early stage in the ETHIC stack, but our formulae make clear how important arousal is—if the body is unable to buy the action-plan, it cannot execute it.[5] Structured breathing (e.g. "box-breathing") is well-known to reduce physiological arousal (e.g. Balban et al., 2023). Stress-inoculation drills that drill breathing like this under stressful circumstances make it likely to be one of the first "purchased" scripts, lowering the overall "neural cash" to buy more action-plans, and thus buying cognitive time for a more calculated appraisal at a higher level (T-level) of the mechanism.

Box 1. What affective circuitry does the literature suggest is relevant for the ETHIC Stack?

See Appendix A (Section 6.0.1) for detail.

Neural substrate

• Subcortical affective networks

• Anterior insula + orbitofrontal cortex (interoceptive hub)

• Cortico-basal ganglia loops (habit machinery)

Core dynamics • Rapid affective tagging of stimuli

• Autonomic arousal that channels attention & primes action

[4] Or, perhaps also -1 if the script is *too* cheap—arousal isn't a frugal mechanism and likes to spend what it has.

[5] Or, if the action plan is being executed, reducing the arousal until the plan is too costly will slow it down or stop it. These kinds of minutiae are difficult to incorporate in a simple illustrative formula though, so I will relegate it to a comment.

• Reinforcement loops that compile habits / "implementation scripts"

Distinctively moral affects Fear, disgust, AND prosocial care

Organising principle (Mandler) Unexpected or goal-blocking events interrupt behaviour → visceral arousal → routinised actions unless counter-programmed.

Intervention levers

1. Stress-inoculation & arousal-control to blunt over-activation
2. Pre-scripted implementation intentions to override default habits

---

4.2 Thought-level schemas

Having addressed the nervous system's initial response, we must then turn to the cognitive and social machinery that receives the output and interprets it. At this level of the mechanism, a great many candidate entities and activities present themselves. Indeed, much of social psychology is concerned with the development of patterns of thinking. It is not immediately obvious to me which have the best claim to a mechanism of moral judgement, so I will simply select, from the overview in Box 2, those that are most convenient for the purpose of this sketch.

It seems to me that, from the candidate theories overviewed in Box 2, we must have at least three broad classes of **entities**: (1) value-generating modules that tie gut feelings to moral foundations; (2) value-editing modules that license us to set inconvenient values aside; and (3) meta-control beliefs about our ability and desire to adhere to our values. The key **activities** these entities engage in would be those described in Section 2: intuitive appraisal, motivated reasoning, and disengagement manoeuvres, all orchestrated around a single organising theme: deliberative cognition is recruited primarily to confirm or defend the gut verdict unless some flag (high uncertainty, high anticipated cost, external accountability) forces a re-appraisal. The **organising principle**, in short, is that of a gatekeeper—allow the intuitive verdict to pass unchanged, or close the gate and *trans-*

form it into a licence to act for better or worse.

4.2.1 A mechanistic sketch

Reminder: I highlight this mechanism because it seems to be the most obvious intervention point, but other plausible mechanics are overviewed in Box 2.

One important feature of this sketch, to make it *moral* and not simply *behavioural* is to explain how gut feelings from the E-level are connected to moral foundations in the first place, or how the moral foundations are crystallised into thought-level schemas from the higher C-level where they are injected into the system. However, for the purpose of a sketch, we will choose another aspect of this system that has already been well described.

The decision to "close the deliberative gate", and force a process of appraisal, is well described by the Expected-Value of Control theory (Frömer et al., 2021; Shenhav et al., 2013):

```
EV_close = P_help * V_help - C_close
```

Where:

- `EV_close` is the expected value of "closing the gate";
- `P_help` is your interpretation of how probable/likely closing the gate will help the situation (which can be most tractably understood as self-efficacy (Frömer et al., 2021)—*can* I deliberate about this);
- `V_help` is the expected value you will get if closing the gate does indeed help (assuming I can, will deliberation be better than not deliberating);
- `C_close` is the cost, in effort, time, or opportunity, of closing the gate (how hard is it to deliberate).

If the expected value of closing the gate is not high enough, we will let the intuition developed at the E-level through. If however, the expected value is high enough, a process of justification will ensue—value-generating and value-editing processes will compete according to the processes overviewed in Box 2—until the expected value is lowered enough to let whatever verdict survived through.

Harmon-Jones and Mills (Harmon-Jones & Mills, 2019) suggest the formula for this justification process in terms of Festinger's Cognitive Dissonance:

> Formally speaking, the magnitude of dissonance equals the number of dissonant cognitions divided by the number of consonant cognitions [cognitions which follow from each other] plus the number of dissonant cognitions [cognitions which oppose each other]. This is referred to as the dissonance ratio.

Or:

```
M_dissonance = D / (C + D)
```

Thus, we will rationalise our cognitions until the magnitude of dissonance is low enough to reduce the expected value of closing the deliberative gate, thus allowing the surving verdict through.

To illustrate, in theatre, the deliberative gate stays open when we'd often prefer to encourage soldiers to shut it (take a "condor moment"). This mechanism implies that we need to (1) raise the impression closing the gate will help,[6] (2) raise the value of closing the gate, and/or (3) lower the cost of closing the gate. A simple way to do the latter is to raise the stakes *a priori*. In theatre, you might prime accountability: tell your soldiers that any use of force will be reviewed face-to-face by the CO and the village elders. Have fire-team leaders repeat the line, and soldiers echo: "reviewed when?", "reviewed tomorrow!" This would increase the value of "closing the gate" in the moment. It wouldn't change soldiers' understanding of how competent they were at deliberating over the need use of force (likelihood closing the gate would help), nor would it change the cost of closing the gate, but it will shift borderline cases over the threshold of deliberation. If more pressure was needed, you might start operating on those other features. Then, of course, you must hope your actions have reduced the cognitive dissonance such that your preferred verdict survives deliberation.

[6] Frömer and colleagues (2021) suggest that this best understood as self-efficacy. In this case, a judgement about our ability to make ethical decisions, perhaps.

Box 2. What does literature suggest happens at the Thought-level schema stage on the ETHIC Stack?

See Appendix (Section 6.0.2) for detail.

Key entities

• Value generators: multiple moral concerns (e.g., Haidt's care, fairness, loyalty, authority, sanctity, liberty – plus Kidder's dilemmas or Ross's prima-facie duties as alternatives).

• Value editors: Bandura's eight moral-disengagement scripts (moral justification, euphemistic labelling, advantageous comparison, minimising consequences, dehumanisation, displacement, diffusion, attribution of blame).

• Meta-control belief: perceived self-efficacy to act ethically.

Core activity

Rapid intuitive appraisal links the affective tag received from Level 1 to a moral foundation, producing an initial "gut" verdict. Gate-keeper circuitry (cognitive control networks) then decides: pass the verdict upward unchanged, or transform / license it via disengagement if the expected benefit of deliberation outweighs its cost.

Organising principle

Gate-keeper weighs conflict, stakes, prediction errors, social surveillance & arousal → recruits deliberation only when worth the effort.

Intervention levers

1. Rehearse intuitions (moral case deliberation) to reduce reliance on the gate-keeper.

2. De-weaponise disengagement (ban euphemisms, force self-blame identification).

3. Retrain attributions (e.g., CBT) & bolster self-efficacy so ethical options become easier to justify.

4.3 Habitat, or immediate situation

Truly, a distractingly-named level, but for the mnemonic, we will persevere.

Situationism is a philosophical lens which notices the overwhelming empirical evidence that variance in human behaviour is usually the function of local, situational triggers, as opposed to some function of "moral character" (see Doris, 2002; Ross & Nisbett, 2011 for review). It is particularly troubling for military ethicists, as military ethics are normally virtue-based ethics—ethics that describe the kind of person one should strive to be. The British Army Values, for example, indicate the soldier should be courageous, disciplined, respectful of others, and so on. If people are more often driven by the situation, then there is a question around how valuable virtue ethics can possibly be (Upton, 2009).

Fortunately, many aspects of humans are relatively stable. For example, personality and IQ would be rather pointless constructs if they weren't fairly stable over time and circumstance. So, we may remain sanguine that character could exist in some stable form among these kinds of psychological qualities (de Bruin et al., 2023).

What we must *not* do is underestimate the importance this level of our model holds. I have made this special note because the activities the entities at this level engage must necessarily affect *other levels*. Since the other levels of this model are merely sketched in brief, what follows may thus may seem unremarkable in context. To be clear, I suspect this level may be the most influential.

Fundamentally, the candidate theories overviewed in Box 3 indicate that **entities** at this level are situations, or situational factors, and the **activities** they engage in are to amplify, dampen, or otherwise interfere with the organisation of other levels. The global **organising principle**, then, is that of a some kind of multiplier.

4.3.1 A mechanistic sketch

Reminder: I highlight this mechanism because it seems to be the most obvious intervention point, but other plausible mechanics are overviewed in Box 3.

The motor-system doesn't weigh every conceivable course of action. Instead, it scans the environment for *affordances*—opportunities for action that the environment presents (Gibson, 2014). When you step onto a lurching subway carriage, the vertical poles don't just *exist* nearby—it invites you to grasp it and stabilise yourself. If the subway carriage is nice and smooth, and a seat is open nearby, you *ignore* the vertical pole, and move toward the seat instead. Two variations in the situation, two different affordances become obvious to you.

The dorsal premotor and parietal "affordance competition hypothesis" (Cisek, 2007; Cisek & Kalaska, 2010) specifies how the brain makes choices like these. Essentially, various action plans are tagged by a situation based on their salience (how close is the pole, how lurchy is the carriage) and utility (will grabbing the pole help, will my arms get tired holding the pole), and a kind of auction ensues (grab the pole, or go sit down instead). Cisek does *not* specify a formal equation for this, but for schematic illustration, we can:

```
AVal_n = (S_n + R_n) - C_n
```

Where:

- `AVal_n` is the value or desirability of a given action plan;
- `S_n` is the perceptual salience of the environment that action plan will operate on—the brightness, speed, size, proximity, urgency and so on;
- `R_n` represents the rewards an action plan might produce under the circumstances—will it reduce the threat, or match your goals;
- `C_n` is the cost of executing the action plan, in time, energy, or exposure, for example.

The action plans which cross some dynamic threshold first become the action plans we

engage in:[7]

```
ASelected = max_n{AVal_n}
```

That is, the selected action plan is the one with the highest action value.

If we assume that two action plans are competing—say, those routinised action plans filtering up from the E-level, and those being generated at the T-level in a moment where the deliberative gate is closed—this mechanism implies that we might think about *increasing* the cost of the automatic plans, and *decreasing* the cost of the alternatives being generated. If a car approaches a vehicle checkpoint too fast, you might have a situation like so:

Competing Plans	Salience	Reward	Cost	Initial AVal
Shoot engine block (automatic)	High (large object, closing quickly)	High (stops threat)	Moderate (routine, but you're not supposed to do that)	High
Shout & gesture stop (deliberative)	Moderate (exposure to drills, ROE back of mind)	High (what you're supposed to do)	**Higher** (must shoulder-tap gunner, expose torso)	Low

The end result is that, even if—in the end—you are *not* going to fire on the vehicle, the automatic action plan is going to start playing out. For as long as the deliberative gate is shut, weapons are coming up and safeties are flicking off. Now, the cost of closing the gate is growing higher—you know you should be trying to *de-escalate* but your action plan is moving in the opposite direction.

[7] This should probably be weighted, or normalised, or perhaps simply probabilistic, not winner-take-all. Or, since Cisek considers mutual inhibitory neural fields, maybe we should include input currents and weightings. However, for a sketch, this will do for illustration.

Rehearsing a stop procedure is an obvious step—fluency reduces the cost dramatically. Also, obstacles placed in front of the firepoint aren't simply tactically advantageous, it also reduces the salience of the vehicle threat by slowing the approach and thus the decision urgency. It might have the added value of increasing the chance of troublesome ricochets, increasing the cost of firing. A quick layout change and a few minutes spent at rehearsal inverts our affordance auction.

As a final note, although it begins to move away even further from Cisek's designs, we might model the weighing of mental alternatives at the T-level along the same principle. If the deliberative gate is shut, and we are now wrestling with the value-editing processes overviewed in Box 2 to justify one decision among many, then we are running a very similar auction. Intervening, to allow the value-based decision we prefer, would use similar logic.

Box 3. What the literature tells us about how context drives (or derails) the ETHIC Stack

See Appendix (Section 6.0.3) for detail.

Key situational factors (Vincent)

• Threat / fear (hostile environment)

• Chronic brutality (normalised violence)

• Sleep-loss, fatigue, time pressure (resource scarcity)

• Authoritative pressure / strict hierarchy

• Ambiguity or poor supervision

Mechanistic impact

– Heightened threat → spikes arousal, rushes E-level habits, magnifies social cues.

– Normalised violence → blunts E-level emotions, leaves room for T-level disengagement.

– Fatigue & time pressure → weaken cognitive control, lower T-level gate-

keeping thresholds, amplify peer / cultural pull.

– Hierarchy → displaces responsibility upward, eases disengagement.

– Ambiguity → raises uncertainty, hands influence to in-group (I) and culture (C).

Candidate theories explaining the link

• Recognition-Primed Decision & Data/Frame Sense-making (Klein)

• Situation Awareness (Endsley)

• Affordance theory (Gibson; neuro updates by Cisek)

• Choice Architecture / "Nudge" as a context-shaping tool

Intervention levers

1. In-the-moment: create a deliberate pause (the "condor moment") or issue clearer orders to relieve ambiguity and restore T-level oversight.

2. Pre-mission: map likely situational pressures to ETHIC levels, then redesign the setting (lighting, ordering of options, wording, team roles, rest cycles, etc.) to favour moral performance.

3. Choice architecture: apply nudges carefully—learn from replication issues; pre-test, measure, and iterate.

4.4 In-group and social dynamics

If the [H]abitat, or situational, level tells us about the particular dangers of context, this level explains why the *same* context affects some groups more than others. It is a particularly good example of mechanism that Glennan (2017) would call a "boundary object": to the T-level below it is a macro-entity (an "in-group") that constrains individual cognition; to the C-level above it is a micro-entity through which culture exercises control. All of my levels are "boundary objects" to some extent for the levels they are stacked between, but this level represents a departure from the more comfortable mechanisms of cognitive and behavioural science, to the socially-defined entities that have less stable boundaries,

even if their interface is still clear.

However, given that we are explaining moral behaviour in the military context, a mechanism with no group-level entities would be necessarily incomplete. More importantly, given that we are talking about the possibility of *intervention*, a mechanism that did not describe cause at the level of the teams commanders would be commanding would have little utility indeed.

The candidate theories overviewed in Box 4 indicate that the **entities** at this level are those that constitute an in-group: social identity representations, the store of group norms, networks of reputation, emotional-contagion loops, and the channels of communication. The **activities** would be the way in which these things allow individuals to accentuate group boundaries, adopt or suppress norms and behaviour, and synchronise affect across the group. The overarching **organising principle** is that of a set of norm pipelines—the norms we most identify with at a given moment will influence the intuitive tags (E-level) and the threshold through which they must pass (T-level), and the pressure on those norms (in those pipes), is a function of identity strength.

4.4.1 A mechanistic sketch

Reminder: I highlight this mechanism because it seems to be the most obvious intervention point, but other plausible mechanics are overviewed in Box 4.

Given the fondness of military ethicists for the Milgram Obedience Experiements, we will derive our mechanism from Reicher and Haslam's (2013) explicit critique of it. The core question Milgram asks is whether the participant should obey their obligations to the experimenter, and the scientific process, and shock a learner for their failings, or should they pay attention to their obligations to a fellow citizen (the learner) and stop. Milgram (Milgram, 1965) explains this with his "agentic state"—a transition from seeing oneself as the author of their acts, to seeing themselves as the agent of, and obedient to, another's will. "Binding factors"—psychological and situational factors that keep a person "bound" to the authority figure—ease or neutralise any moral strain.

Reicher and Haslam (2013) point out that Milgram's explanation ignores the learner:

> If one were to take it seriously, it would suggest that the participants are serene, that the cries of the learner have no hold on them and that they are only interested in doing the experimenter's bidding. In fact, psychologically, it writes the learner out of the picture. (p. 121)

They suggest, instead, "The Engaged Follower" better explains the situation. If the "follower" identifies with the scientific enterprise (i.e. the experimenter who *represents* it) more than they identify with the plight of the learner, they will shock the learner. As soon as the identification flips, they no longer comply. For example, Reicher and Haslam point out that:

> Milgram scripted a sequence of four prompts (prods) for the experimenter to use if and when participants proved reluctant to continue. Three of these are either requests or justifications linked to the scientific importance of the participants' acts: "Please continue", "The experiment requires you to continue", "It is absolutely essential that you continue". Only the fourth and final prompt – "You have no other choice, you must go on" – constitutes a direct order. The interesting finding is that whenever Prompt 4, the order, is used, people typically refuse to continue (p. 123)

The authors suggest that this means the participant is not being *obedient*, in the sense of blindly or "agentically" following orders. They are being *compliant* because they feel they *ought* to comply. As soon as the identification is broken—by the jarring order to continue—they no longer feel that way, identifying with their fellow citizen and complying with *those* obligations instead.

So, we might consider the calculus of an engaged follower thus: an engaged follower will comply with the aims of the representative of an identity category to the extent that the representative elicits binding factors, and the extent to which those binding factors outweigh those of other available category representatives:

```
IWilco = max { all of the (R * B) for each I }
```

Where the likelihood of complying with the aims of an identity (`IWilco`) is the highest

value of the binding factors (B) multiplied by the representatives eliciting those factors (R) for each identity (I).

This presents three levers:

1. Increasing the representatives of an identity category. We know in later studies, when a second 'participant' sided with the learner the original participant would no longer deliver shocks.
2. Changing the perception of how much the representative of an identity category represents the category. When the experimenter switched to the less scientific order, the participant no stopped shocking the learner. Equally, later, Milgram has two *arguing* experimenters, and again compliance dropped (Milgram & Gudehus, 1974).
3. Manipulate the binding factors—psychological and moral factors that ease moral strain. In this case, the experiment was framed deliberately from "something socially pernicious to something socially progressive" (Reicher & Haslam, 2013, p. 124).

Box 4. What does the literature say about in-group & social-dynamic influences on the ETHIC Stack

See Appendix (Section 6.0.4) for detail.

Key entities

• Social identities & self-categorisation (Tajfel, Turner)

• Optimal distinctiveness motives (Brewer)

• Identity leaders / influential nodes ("stars") shaping norms

• Prestige economies that reward conformity & police deviance

• Mechanisms of emotional contagion within the group

Core activities

1. Identity adoption → stronger identification amplifies group norms.
2. Norm definition by identity leaders spreads through prestige signals.

3. Emotional contagion rapidly aligns arousal & moral stance across members.

Organising principle

The more a situation highlights a valued in-group identity, the more its norms steer E- and T-level processing; identity leaders and prestige cues accelerate this alignment, while devaluing out-groups can strip them of moral standing.

Intervention levers

• In-the-moment: prime a constructive nested identity (e.g., "platoon ethos" over a corrosive sub-clique); spotlight ethical norms.

• Ongoing: curate and rehearse group norms; train identity leaders to model them; tether status & reputation to ethical as well as operational success.

• Systemic: engineer communication channels—rotate cliques, foster cross-talk, install anonymous "red-card" systems—to check corrosive norms and encourage ethical voice.

4.5 Cultural and institutional scaffolding

The last level is perhaps the most critical, even if it is the least manipulable. The cultural and institutional scaffolding is the place where moral content is *injected into the system*. Outside of group norms, which are normative in their own way, the levels E-I simply describe processes that influence the transformation of moral content into behaviour. They are descriptions of *what is* not *what should be*. This level, instead, supplies the standards, narratives, and reward structures that define which actions are counted as right and wrong.

The **entities** at this level, inspired by the candidate theories overviewed in Box 5, are the

large-scale normative frameworks[8] that surround a fighting force, and the broader cultural schemas in which a unit is embedded. The **activities** these entities engage in are myth-making, codification, ceremonial exemplification, framing, and explicit systems of sanction and reward. The overarching **organisational principle** is that of the top-down constraint—the high-level priors which biases the operation of each lower level.

4.5.1 A mechanistic sketch

Reminder: I highlight this mechanism because it seems to be the most obvious intervention point, but other plausible mechanics are overviewed in Box 5.

Since I have little business suggesting literature-derived mechanism from the tier I know least about, I will instead derive my mechanism from Bayes Theorum—a simple formula used for calculating probabilites that is explicitly biased by high-level priors (and indeed is where that term comes from)—and has been used to explore belief in philosophical literature before (Lin, 2024).

We might imagine each solder as having a "moral warrant" score for their actions that is continually updated. The cultural level supplies the *starting point* for that running total: a *normative prior*. If the prior is strong enough, it is unimportant how much nudging from the lower levels—E-level visceral tags; T-level reasoning; H-level affordances; and I-level norms—occurs, the "moral warrant" remains stable. For illustration, we could model with a deliberately spare Bayesian-style rule:

```
W_post(A) = t * P_C (A = 'acceptable')  + each { w_k * E_k }
```

We imagine that every action (`A`) has a moral warrant score (`W_post(A)`, the "posterior"). This is a combination of several things. First, our understanding of how acceptable that action is in our culture (`P_C(A = 'acceptable')`, the "prior"). This is multiplied by how strongly we think our culture penalises unacceptable behaviour (`t`, a tight/loose culture (Gelfand et al., 2011) modifier). But also, we must add all the evidence we're getting

[8] Normative frameworks are structured sets of values, principles, and rules that articulate how people "ought" to, or should behave. The British Army Values and Standards are an example of a normative framework. A member of the British Army *should be* courageous, disciplined, respectful of others, and so on. They *ought to behave* lawfully, appropriately, and totally professionally.

in the moment for how likely it is we're going to be punished, coming from the other levels of the ETHIC stack—each of the `E_k`, or evidences from each `k` level, modified by how *strong* the evidence is—the "weight", `w_k`.

So, doctrine tells us we must *never* shoot a wounded detainee. The prior (`P_C('shoot' = 'acceptable)`) is very low, and so is our moral warrant score (`W_post('shoot')`). Courts-martial for this sort of thing can be swift and public, so we know that the military is *tight* on this and our gain term (`t`) is big—our score is made even *lower*. Suddenly, the detainee twitches and a squad-mate yells "weapon!" Adrenaline floods the system and our E-level starts to work—there's some new evidence (`E_E-level`) in the system. Not enough to amount to much, though. The moral warrant score isn't changed enough to make you shoot the detainee.

Box 5. What does the literature suggest about Culture-level influences – the normative backdrop of the ETHIC Stack

See Appendix (Section 6.0.5) for detail.

Key entities & theories

• Tight vs. loose cultural norms (Gelfand)

• Collective memory & heritage (Ashworth)

• Institutional logics / organisational fields (DiMaggio, Thornton)

• External legal-normative regimes (e.g., IHL / LOAC)

Core activity

Culture injects overarching rule-sets and narratives that pre-bias all lower levels toward (or away from) particular moral stances; these frameworks diffuse through institutional channels and become "the way things are done."

Organising principle

The density and strength of shared norms (tight vs loose) dictate tolerance for deviance; institutional stories & legal codes supply the moral content that individual leaders must interpret and operationalise.

Intervention levers (limited reach)

• Strategic: doctrine & policy revision; institutional storytelling; heritage framing.

• Tactical: emphasise or down-play existing cultural narratives to support immediate ethical aims (leverage lower ETHIC levels to translate broad norms into action).

NB: For most leaders, culture is a background constraint rather than a direct control knob—work with it more than on it.

---

5 Conclusion

In this paper, I have argued that ethical competence in combat emerges from a chain of nested mechanisms—-visceral, cognitive, situational, social, and institutional. By naming those levels and the levers attached to each, the ETHIC Stack turns abstract value talk into an engineering diagram commanders can actually manipulate.

To be clear, the purpose of this mechanistic sketch is not to *replace* the many ethical decision-making models that already circulate within PME. The ETHIC Stack is offered as the causal plumbing that services them—informing which mechanistic levers might best fulfil their prescriptions, and thus addressing their chronic defects. My hope is that, given mechanisms worked to the level of their formulaic dynamics, ethicists may test and refine the Stack by plugging their models and value-sets into its levels, and vice versa.

In closing, let me illustrate how the Stack (1) accommodates and augments what is already on offer, (2) sharpens current trends in ethical education, and (3) highlighting where the framework is weakest, and in most need of refinement. On this last point, the most critical vulnerability is the need for explicit cultural tailoring. Without the explicit injection, by the user, of normative frameworks at the C-level, the model remains simply an incomplete account of human behaviour, rather than a mechanistic account of *ethical* behaviour.

5.1 Augmenting existing decision-aides

First, let me demonstrate how the model sharpens existing models. As detailed in Section 2 and Section 3, Kem (2006) supplies three ethical *lenses* (rules, consequences, and virtues) through which to explore behaviour, but provides no lever for decising which lens should win when they conflict. In the ETHIC Stack, those lenses sit naturally at the [T]hought-level: they are content-rich schemas among which the gatekeeper must choose. The problem is then refamed as a gatekeeping one—which valuation schema reaches threshold first, and can a commander raise or lower that threshold by working a handle at the other four levers? The commander might recognise a hot limbic surge (E) that flagged the event as a moral risk, and understand that without intervention, it will move quickly past the deliberative gate (T)—automatically tethered ethical lens in tow. By priming a different social norm (I) the commander can lower cost of closing the gate, while also stripping euphemisms from their pattern of thought that obscure problematically biased rationalisations *at* the deliberative gate (T).

The process just described conveniently[9] maps onto another, more recent model of ethical decision-making—Ordiway's (2022) Moral Terrain Coaching process. Ordiway, who inspired this particular project of mine with his work, has coaches help those they're coaching recognise visceral and situational factors which contribute at the E-level to moral judgement; recognise the ethical conflicts through the lens of Ross' (2007) *prima facie* duties and Bandura's (2011) disengagement manoeuvres occurring at the T-level; and apply a social norm (I-level) by asking "what would someone I look up to think"; all of which are used to improve E-level intuitions and T-level deliberative processing in the future.[10]

Similarly, Sandhursts' own S-CALM model (Vincent, 2022) is made more complete. Rather than a list of thirteen biases and five situational risks, it becomes a list of moral terrain features we can now map. The situational factors belong at the H-level, and am-

[9] It was not convenient. I did it on purpose.

[10] Indeed, Ordiway has evidence to suggest this is true, though the publication of the findings is not yet available to me.

plify the arousal of the E-level, making it less likely the deliberative gate at the T-level will close. Most of the common behaviours are classic I-level phenomena—norm pipelines that push or pull behaviour once group identity is salient. Others sit at the T-level as value editing entities that imply moral disengagement manoeuvres when the H-level turns up the heat. We now have the tools to begin exploring why and how these features coalesce to produce unethical behaviour in military contexts, which will then make which features of his three lines of moral questioning are most likely to improve the situation clear.

A final model, perhaps the most mechanistic I have encountered in this domain, is Messervey and colleages' (2021) Revised Defence Moral Decision-Making Model. It is a laudably complete account of the E- and T-levels of Stack. The H-, I-, and C- levels stand to add greater fidelity on their mechanisms of interest, should they wish to test their empirical claims further.

5.2 Sharpening recent PME trends

PME syllabi in the UK, US, Canada, and elsewhere have already begun to move past rule memorisation and toward the kinds of interventions we see here. Commonly, moral rehearsals, ethical readiness drills, stress-inoculation, and cultural and identity awareness form part of their recommendations (e.g. Karssing, 2023; Messervey et al., 2023; Ordiway, 2022). The ETHIC Stack provides a coherent skeleton on which to hand these disparate efforts:

- arousal-control lanes and ethical drills are E-level interventions that improve the intuitions and habitual patterns that arise from affective tagging;
- moral case deliberation and ethical wargaming are T-level interventions that sensitise individuals to those intuitions and refine the activity of the deliberative gate toward moral—and not merely convenient—licence to act;
- both of these classes of intervention probe the situational factors which sit at the H-level;
- prestige-economy designs (recognition and awards) manipulate the I-level;
- after-action reviews and myth-making enterprises increase C-level influence;

and so on. Mapping training onto distinct levels thus prevents the primary defect in ethical PME diagnosed in Section 2 of this article—over-weighting deliberation—while also allowing PME planners audit where their programme remains thin.

5.3 Obvious vulnerabilities, limitations, and opportunities for refinement

The primary vulnerability of the model is that it is *not* normative. It is a model of *what is*, not *what ought to be*. The purpose of the model is to explain how individuals act based on moral content, but contains none. As such, it must be *populated with local norms* before the mechanism can tell us anything about right or wrong (though c.f. Cushman & Young, 2011). Again, this is the value of *combining* it with other models—Kem's (2006) ethical lenses for example, or the British Army Values and Standards.

Equally critically, without this feature—the injection of moral content—the model simply lists generic components of human behaviour which have no particular reason to hang together in the manner I have described. What makes the ETHIC Stack a *moral* mechanism rather than a mechanism of behaviour writ large is that, at each tier, we must clearly single out *those entities and activities specifically recruited when an event is coded as morally significant*:

- the **E**-level valence system must be tuned to norm violation or fulfilment;
- the **T**-level gatekeeping must be tuned to culturally-parameterised content-rules and their adjudication;
- the **H**-level must be tuned to environmental signals that toggle our moral circuitry;
- the **I**-level identification processes must be tied directly to a moral standard; and
- the **C**-level must frame certain acts as transgressions or exemplars, and attach real penalties or honours.

In the absence of these constraints, we would simply be providing a strikingly incomplete "model of everything".

More generally, the Stack is, at best, a "how-plausibly" sketch of moral decision-making.

Four more liabilities are particularly apparent:

1. My framework is agnostic about which values flow in at the C-level and crystallise at T-level. It is likely that some moral content-rules travel better than others. Perhaps, as those like Ross (2007) would have it, some are universal, which would add an additional injection-point for moral content at the E-, or perhaps T-level. Exploring the norm-injection mechanics empirically to flesh out the most critical, and incompletely described C-level seems prudent. In particular, I am only really considering an anglosphere military context, and even then I only lightly consider norm-injection of the relevant normative frameworks. Considering from other cultural perspectives seems important.
2. Two more, related, underspecifications stand out. The first is that of the situational H-level. Situational factors are perhaps the most influential on human behaviour. Vincent has identified those which seem most correlated to unethical behaviour in the military context. Empirical treatment of these, within the ETHIC framework seems like it should be prioritised. More generally, beyond the H-level which necessitates it, I otherwise left almost entirely undiscussed the *inter-level* mechanics. How might platoon norms (I) sediment into regimental myth (C) for example? Or dialogue around disengagement (T) rewire affective tags (E)? For the purpose of diagnosing which handles have the most utility both in- and out-of-the-moment, a programme seeking to describe these would be important.
3. The brain is multi- rather than dual-process, as described in Section 2 of this paper. The current E-/T-level partition borrows from dual-process theories, and this will almost certainly require attention as the idea of a "society of mind" (Mercier & Sperber, 2017; Minsky, 1986) begins to once again come into fashion.
4. Why five levels, and not four or six? My selection criteria was determined when the dominant manipulable organisations shifted domain—neural, cognitive, situational, relational, and cultural. However, one could make a case for further splits—the three classes in the T-level, for example. Equally one could make a case for combining levels—situational triggers and in-group pressures seem like strata that could occupy a single, "context" level. The ETHIC acronym is cute for the

purpose of a memorable sketch, but should by no means constrain the model if the levels must be adjusted.[11]

5.4 Final word

The ethical application of extreme violence is the business of warfighting. It's not simply a humanistic imperative, but the determinant of post-conflict legitimacy. It emerges from a stack of nested mechanisms which begin in the gut, pass through a largely misunderstood process of "reason", are modulated by the moment, shaped by our tribes, and ultimately authorised by culture. By exposing these mechanisms, and more crucially, flagging where commanders can intervene upon them, the ETHIC Stack turns abstract value statements into actionable engineering diagrams. Imperfect and incomplete, the model nevertheless provides what the overwhelming majority of ethical models lack: a map of the causal, moral terrain upon which we must fight.

[11] And no, for those asking the reasonable question, the acronym did not produce the levels. I certainly would not have chosen 'early emotional' or 'habitat', which if anything obscure the purpose of the level. You may thank the *post hoc* application of AI for that happy product.

6 Appendix A: Candidate Mechanism Theories and Plausible Interventions

In an effort to guide development, I have corralled some of the relevant literature that could describe the entities, activities, and organisation of the mechanism at each level of the ETHIC Stack, as well as what that might mean for intervention.

6.0.1 Early, affective circuitry

Candidate neural substrate in which to house the E-level entities would almost certainly include the subcortical affective networks of the brain (Dalgleish, 2004; Kandel et al., 2013), the interoceptive integrative properties of the anterior insula and adjoining orbitofrontal cortex (Craig, 2009), and the cortico-basal loops which appear to form the basis of our habit machinery (Graybiel, 2008).

The activities are less easy to house comfortably in theory, but we can make an attempt. The mere exposure effect appears to capture precisely the kind of pre-cognitive phenomena I am interested in, and the Representation-Matching model of Montoya and colleagues (2017) seems an excellent place to start, not least because their meta-analytic findings indicate "people, without conscious processing or (mis)attributional assessments, come to evaluate well-learned stimuli as 'correct' and 'how things should be' " (p. 476). We might also consider something like LeDoux's (LeDoux, 2012) or Damasio's (Dunn et al., 2006) account of the rapid affective tagging of objects which bias our appraisals, though the evidence in support of them is far from settled. The autonomic arousal that funnels attention toward the tagged object and prepares stereotypical motor programmes would also play a role (e.g. Lang, 1995). And reinforcement loops that compile habits, or scripted "implementation intentions" would complete a preliminary set (e.g. Gollwitzer & Sheeran, 2006).

What seems critical, however, is to isolate *moral* affect in particular. Threat and fear are well documented, but insufficient. Disgust also seems an early affective valence (Calder

et al., 2001). But morally relevant emotions aren't solely negative. Panksepp (Panksepp, 2005) for example includes care as an early valence in both humans and animals, and the aforementioned mere exposure effect is approach oriented (liking/familiarity) in nature, both of which present interesting targets.

The global organising principle is easier to land. George Mandler's (Mandler, 1980; see Barrett & Russell, 2014 for modern developments) foundational work on the generation of emotion seems an entirely appropriate fundament. A basis for much of the social psychology of emotion, and derived from "an extensive, but discontinuous, line of reasoning", Mandler noted that the:

> common insight is that many if not all emotional states arise out of the interruption of ongoing psychological events, out of the conflicts and the discrepancies among them, or from the frustrations of actions.

Thus, an unexpected event—some norm violation or the elicitation of moral disgust perhaps—or a goal-thwarting frustration—an obstacle to norm fulfilment, for example—interrupts ongoing behaviour. Visceral signals respond by mobilising autonomic energy and then hand off to higher levels a problem already helpful tagged with a valence—threat, reward, or uncertainty, for example. Under high arousal, routinised action plans are preferentially released unless there is some kind of counter-programming in place.

6.0.1.1 Plausible Interventions Our first level therefore presents at least two obvious targets for intervention before these early, affective biases cascade upwards. Firstly, stress-inoculation drills and arousal-control training would blunt over-activation of the mechanism. Secondly, pre-situational "implementation intention" scripts might work as counter-programs to help override the routinised behaviour the hot affective tags our mechanism would otherwise produce. Each seem like plausible ways to 'wiggle' the mechanism in order to change the outcome.

6.0.2 Thought level schemas

With regard to value-generating entities, given what Haidt has contributed to the domain already, it seems appropriate to begin there. A large and diverse empirical record supports Haidt's basic claim that multiple, partially independent moral concerns structure everyday intuitive judgement—care, fairness, loyalty, authority, sanctity, and liberty (Haidt, 2001, 2012). However, other perspectives on the same class of entities exist. We have already spoken of Kidder's (1995) common ethical dilemmas—truth versus loyalty, justice versus mercy, short-term versus long-term, and individual versus community. Ordiway (2022), the military ethicist who inspired this project of mine, prefers Ross' (Ross, 2007) *prima facie* duties: an urge to keep our promises, an urge to make amends, an urge to return favours, an urge to improve the lot of others and ourselves, an urge to do no harm, and an urge to distribute pleasure according to merit.

These entities, whichever framework one chooses, engage in the activity of intuitive appraisal—the rapid matching of the affective tag passed from the previous level to a moral foundation which determines a gut verdict as to the 'rightness' of a thing. This activity could be couched in any almost any cognitive appraisal theory (e.g. Gigerenzer & Gaissmaier, 2011; Kelley & Michela, 1980; Lazarus, 1991)—anything which connects a feeling to cognitive content.

Ordiway (2022) also recommends a candidate for the second class of entities—the value-editing modules. For this, he suggests Bandura's catalogue of moral disengagement mechanisms (Bandura, 2011). These are linguistic or cognitive reframing manoeuvres that—as an activity—dampen our self-sanctions. Bandura details eight, and each is worth mentioning, as each comes with a characteristic flavour:

1. Moral Justifications allow us to reframe an act as as serving a worthy or higher purpose ("we had to torture him to save lives").
2. Euphemistic Labelling is where we swap harsh descriptors for sanitised ones ("enhanced interrogation", "collateral damage").
3. Advantageous Comparisons contrast a contemplated act with something far worse

("at least we're not torturing him like the last unit did").

4. Distortion/Minimisation of Consequences is the process of downplaying harms ("he'll be fine in a couple of days").
5. Dehumanisation is the process of denying a victim full moral status ("they're savages; they don't feel pain the way we do").
6. Then three that centre on the re-attribution of responsibility:
    - Displacement of Responsibility allows us to attribute agency to another figure ("I was just following orders").
    - Diffusion of Responsibility is where we spread agency across a group, rather than attribute it to an individual ("the whole platoon agreed this was necessary").
    - Attribution of Blame is where we cast the victim as responsible for the harm ("he brought it on himself by running").

Bandura also has a candidate for our final class of entity—meta-control beliefs about our ability to adhere to our values. Bandura's notion of self-efficacy (Bandura, 1982) describes our beliefs around our capacity to do something—in this case, enact an ethical alternative. Activity-wise, low self-efficacy may make disengagement scripts more attractive, or encourage us to defer to a higher level in the ETHIC Stack for input, where high self-efficacy would do the opposite. I am not wedded to Bandura for this, I should point out, he simply served as a convenient segue. Any theory that attends to the role of uncertainty in decision-making seems as though it might service the same explanatory gap.

Lastly, the overarching organising principle—that of a gatekeeper which passes the intuition onwards unchanged, or transforms it—seems to have an obvious overlap with the literature on cognitive control.[12] Converging work (e.g. Seeley, 2019; Shenhav et al., 2013) appears to demonstrate that certain networked regions of the brain synthesise precisely the kinds of red flags we are speaking to—response conflict, outcome stakes, errors in prediction, social surveillance, and visceral arousal—into a single quantity: is the cost of recruiting deliberation justified by the expected benefit? When that quan-

[12] Though equally, this might only be true because this is my speciality.

tity crosses threshold, deliberative resources are engaged, often to actively licence the intuition through disengagement; when it does not, the intuitive verdict passes upward unedited.

6.0.2.1 Plausible Interventions

How might one intervene? This architecture isn't intuitive. We almost want to avoid activating the gatekeeper, for fear of licensing moral disengagement. It would seem to imply, rather, that we must rehearse our intuitions—practice resolving value conflicts for example, or otherwise engage in moral case deliberation.

We can however, work to make the deliberative action less risky. We might de-weaponise disengagement mechanisms—avoiding or outright banning euphemisms for example, or requiring us to identify our own blame before we are permitted to recognise the blame of others. We could also retrain our attributions—cognitive-behavioural therapy, for example, is designed precisely for such an enterprise. To make the task more tractable still, we might observe, as Mercier and Sperber (2017, introduction) did:

6.0.3 Habitat, or situational factors

Vincent's (Vincent, 2022) situational factors seem like an excellent starting point given that they are derived from military case studies. Each of his factors implies an associated activity:

- Danger, threat, and fear (hostile environment/emotional compromise) spike arousal, hastening the E-level tagging and release of habitual scripts, while also sharpening the kinds of social effects I am yet to describe at the I-level.
- Exposure to chronic brutality (normalised violence) has the opposite effect, dampening E-level affective tagging, while also providing room to engage in disengagement manoeuvres at the T-level.
- Sleep deprivation, fatigue, and time pressure (lack of resources) impedes cognitive control, and thus lowers the thresholds for gate-keeping at the T-level, while also making one more susceptible to E-level as well as in-group (I) and cultural (C) level influences (again, as yet undescribed).

- Authoritative pressure and hierarchical obedience (weak leadership) displaces responsibility upwards, perhaps emhapsising I-level in-group influences and certainly encouraging T-level licensing of behaviour.
- Ambiguity (lack of supervision) increases T-level uncertainty, enhancing the influences of I- and C-level influences.

So on, and so forth. Each of these are uncontroversial claims, sketched as they are in the broadest possible terms. A situational factor acts to enhance or inhibit the mechanism at one or more levels, with a corollary effect on the action of the mechanism at the remaining levels. The global organisational effect at this level thus has the potential to be the most pervasive.

Some likely candidates for exploring this more empirically seem obvious. We might start with two- (Dror, 2017; Zillmann, 2008) and three-factor theories (Russell & Mehrabian, 1977) of sympathetic emotional transfer from one event to another. Alternatively, Klein's (1993) Recognition-Primed Decision-Making, or his later advancement in his Data/Frame Model of Sensemaking (2007); Endsley's Situation Awareness model (2012); and Gibson's (2014) Theory of Affordances, as well as modern advances into neuroscience (e.g. Cisek, 2007) all describe the ways in which contextual situations select for particular T-level gating thresholds—they specify what cognitive schemas should become active.

One particularly likely candidate for *intervention* might be the behavioural economist project around Choice Architecture (Hansen, 2016; Johnson et al., 2012). This is the idea, made famous by Thaler and Sunstein's pop-science book "Nudge" (2008), that the small, often incidental details of a decision environment (defaults options or the order of them, framing and visual salience, etc) can be deliberately arranged so as to steer people toward welfare-enhancing choices without restricting the set of options available to them. While the empirical evidence in support of the project have recently come into a troubling degree of contention (Maier et al., 2022; Osman et al., 2020), it seems a good candidate program to explore how the decision environment can steer morally-enhancing choices too, assuming that we also adopt the lessons of the programmes failings.

6.0.3.1 Likely Interventions Situational factors are those we are least able to change "in the moment". Sleep deprivation, or time pressure are rarely things we can address once they are occurring. However, a rough understanding of how common situations interact with levels of the ETHIC Stack may permit insight into how to free up a level with more perceived utility. For example, the "condor moment" so revered by Officer Cadets—stopping to breathe and think—is likely to alleviate constraints on T-level gating, making the chances that an intuition is subject to re-evaluative pressure more likely. Or, improving the quality of orders would reduce uncertainty, allowing team-members to take greater ownership over their decision-making (again, referring to the as yet undescribed [I]n-group and [C]ultural levels).

Situational factors are, however, most easily addressed *outside* the moment. Building in planning time to walk through the more obvious situational pressures, mapping them to their ETHIC Stack effects, and deciding on some small number of interventions to achieve the desired effect seems plausible indeed. Of course, this depends on determining more clearly what situational factors hold most sway over the other ETHIC levels, which will require some empirical exploration. In particular, given the spectacular failings of the "nudge" programme (Maier et al., 2022; Osman et al., 2020), particular care must be paid to research in this domain.

> The solitary use of reason has two typical outcomes. When the reasoner starts with a strong opinion, the reasons that come to her mind tend all to support this opinion. She is unlikely, then, to change her mind; she might even become overconfident and develop stronger opinions. But sometimes a reasoner starts with no strong opinion, or with conflicting views. In this case, reason will drive her toward whatever choice happens to be easier to justify, and this sometimes won't be the best choice.

On this account, we simply must do whatever will make our desired outcomes easier to justify.

6.0.4 In-group and social dynamics

The most critical candidate theories, both to describe many of the entities and activities, but also to describe the organising principle, are those that surround the ways in which we adopt our social identities. So Social Identity and Self Categorisation Theory (Tajfel & Turner, 1979) combined with Brewer's (Brewer, 2009) Optimal Distinctiveness indicate that when individuals can see the ways a group represents who they are, or wish to be, in comparison to other, similar groups, they will accentuate their membership to that group in their behaviour. Thus, the greater the identification, the more group norms are likely to influence our behaviour.

Two more forces also seem at play in the influence of group norms. The first might be best described by Reicher and Haslam's work on "identity leaders" (2005; Reicher & Haslam, 2013) —individuals with particular influence over norm definition. This force might equally be well described by the concept of 'influential nodes' within human systems literature, coined, I believe, with Moreno's (1934, Chapter 13) "stars".

Alongside our influential identity shapers, we might also consider the role of prestige economies in policing conformity, but also in stripping out-groups of moral standing (e.g. Blader & Tyler, 2002; Clayton & Opotow, 2003; Opotow, 1990). There is an important psychological difference, in in-group comparison, between noting the virtues of your own group and the failings of another.

Grossman, in *On Killing* (2014), does not organise his content around these ideas, but reading his chapters will yield a fairly discrete package that resemble these ingredients: authoratative sanctions, group absolution, dehumanisation, moral displacement, absence of countervailing norms, etc.

Lastly, we must consider the way in which a group can rapidly spread affect among the individuals. At a minimum, we should consider Hatfield and colleagues' (1993) concept of Emotional Contagion—the process of 'catching' the feelings of others through the synchronisation of expressions, cues, and other behaviours. Also relevant would be Haslam and Reicher's work on the role of identity in mobilising groups (Reicher & Haslam,

2012). It also feels as though more could be made of the literature on inter-group communication and the vectors for these things, but for now this will suffice.

6.0.4.1 Likely Interventions This [I]n-group level appears, *prima facie*, obvious as the lever with greatest mechanical purchase. Emphasising a single norm has the potential to re-route innumerable T-level gatekeeping decisions and generate affiliated E-level affect for as long as the norm remains dominant. Ignore the I-level, and even the most rigorous E- and T-level training could be undone by a particularly corrosive platoon climate.

Social identities are naturalistically nested. The fireteam sits within the section sits within the platoon sits within the company, and so on. Intervening in the moment could be a simple as priming one of these nested groups over the other. Otherwise, crafting the context that makes the right group norms salient could be as simple as emphasising the positive distinctions that are reflected by your group as opposed to some deficient outgroup.

Outside the moment, one should explicitly curate group norms and train identity leadership to steward those norms throughout the group. This means training other leaders and influential team-members to represent those norms and advance them, as well as helping them craft contexts that make those norms salient. We would also want to harness status and reputation. Tying reputational credit, for example, to ethical as well as operational or administrative successes seems important. Lastly, we would want to manage our channels for intergroup communication. Rotating cliques, fostering cross-talk and installing anonymous 'red card' systems are all top-of-the-head interventions that seem quite plausible.

6.0.5 Cultural and institutional scaffolding

Of all the levels, I have the least to offer here, and we have the most to gain from existing ethical decision-making literature. For example, literature on legal-normative regimes like International Humanitarian Law and the Law Of Armed Conflict is commonly touched upon by more knowledgeable military ethicists than myself, and should certainly be considered here.

I will briefly suggest some candidates which spring to mind, however. Obviously the work on Social Capital bridges small and large collective action (e.g. Coleman, 1988; Patulny & Lind Haase Svendsen, 2007; Portes, 1998; Putnam, 1995). Gelfand and colleagues' (2011) work on "tight" and "loose" cultures describe the multi-level factors that contribute to the number and strength of norms within a culture, and the extent to which that culture tolerates deviance. As a means of filling in my tentative sketch, it seems to pose a great deal of promise. I am also aware of Ashworth's (e.g. Ashworth, 2011) work on collective memory and heritage in shaping cultural identity. Lastly, I have attempted to puzzle through material on Institutional Logics and Organisational Fields (e.g. DiMaggio et al., 1983; Thornton & Ocasio, 2008), which I am told speaks to how rule systems diffuse across collectives, making certain practices more or less natural.

Those mentioned, and given that this level is the least manipulable by the average leader, it seems least worthy of greater attention in this early stage of the ETHIC Stack sketch. The critical feature of the level is to inject the normative frameworks which make the remaining levels relevant to study.

6.0.5.1 Likely Interventions The C-level does not offer much mechanistic purchase. Doctrinal changes perhaps, and policy interventions, if they are within reach. Narrative framing and institutional storytelling are also aspects of the C-level that leaders may be asked to participate in.

However, unlike the previous levels, the C-level is not manipulable for the average commander. It injects moral content into the system which biases the rest of the system. The role of the commander is to then moderate or emphasise this content using lower levels of the stack to achieve more local moral ends.

References

Ashworth, G. J. (2011). Heritage in Ritual and Identity. In *Ritual, Heritage and Identity*. Routledge India.

Balban, M. Y., Neri, E., Kogon, M. M., Weed, L., Nouriani, B., Jo, B., Holl, G., Zeitzer, J. M., Spiegel, D., & Huberman, A. D. (2023). Brief structured respiration practices enhance mood and reduce physiological arousal. *Cell Reports Medicine*, *4*(1). https://doi.org/10.1016/j.xcrm.2022.100895

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Bandura, A. (2011). Moral Disengagement. In *The Encyclopedia of Peace Psychology*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470672532.wbepp165

Barrett, L. F., & Russell, J. A. (2014). *The psychological construction of emotion*. Guilford Publications.

Bechtel, W. (2007). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Psychology Press.

Beck, A. T. (1979). *Cognitive therapy and the emotional disorders*. Penguin.

Blader, S. L., & Tyler, T. R. (2002). Justice and empathy: What motivates people to help others? In *The justice motive in everyday life* (pp. 226–250). Cambridge University Press. https://doi.org/10.1017/CBO9780511499975.013

Brand, C. (2016). Dimensions of Moral Intuitions – Metaethics, Epistemology and Moral Psychology. In C. Brand (Ed.), *Dual-process theories in moral psychology: Interdisciplinary Approaches to Theoretical, Empirical and Practical Considerations* (pp. 19–40). Springer VS.

Brewer, M. B. (2009). Motivations underlying ingroup identification: Optimal distinctiveness and beyond. In *Intergroup Relations* (pp. 22–41). Psychology Press.

Calder, A. J., Lawrence, A. D., & Young, A. W. (2001). Neuropsychology of fear and loathing. *Nature Reviews Neuroscience*, *2*(5), 352–363. https://doi.org/10.1038/35072584

Centre for Army Leadership (CAL). (2021). *Army Leadership Doctrine - AC 72029*. British Ministry of Defence.

Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1585–1599. https://doi.org/10.1098/rstb.2007.2054

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, *33*(July), 269–298. https://doi.org/10.1146/annurev.neuro.051508.135409

Clayton, S., & Opotow, S. (2003). Justice and Identity: Changing Perspectives on What Is Fair. *Personality and Social Psychology Review*, *7*(4), 298–310. https://doi.org/10.1207/S15327957PSPR0704_03

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, S95–S120.

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70.

Craver, C. F. (2007). *Explaining the brain : Mechanisms and the mosaic unity of neuroscience / Carl F. Craver*. Oxford : Clarendon, c2007.

Crayne, M. R. (2025). Our Most Powerful Weapon: The Army Ethic. *Harding Paper*, *25*(1).

Cushman, F., & Young, L. (2011). Patterns of Moral Judgment Derive From Nonmoral Psychological Representations. *Cognitive Science*, *35*(6), 1052–1075. https://doi.org/10.1111/j.1551-6709.2010.01167.x

Cushman, F., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm. *Psychological Science*, *17*(12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x

Dalgleish, T. (2004). The emotional brain. *Nature Reviews Neuroscience*, *5*(7), 583–589. https://doi.org/10.1038/nrn1432

Davis, M. S. (1971). That's interesting!: Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences*, *1*(2), 309–344. https://doi.org/10.1177/004839317100100211

de Bruin, B., Zaal, R., & Jeurissen, R. (2023). Pitting Virtue Ethics Against Situationism: An Empirical Argument for Virtue. *Ethical Theory and Moral Practice*, *26*(3), 463–479. https://doi.org/10.1007/s10677-023-10381-5

Deakin, S. (2014). Leadership: Proceedings Symposium Held Royal Military Academy Sandhurst, 33. *Leadership: Proceedings Symposium Held Royal Military Academy Sandhurst,* 33.

DiMaggio, P. J., Powell, W. W., et al. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. Cambridge University Press.

Dror, O. E. (2017). Deconstructing the "two factors": The historical origins of the schachter–singer theory of emotions. *Emotion Review*, *9*(1), 7–16. https://doi.org/10.1177/1754073916639663

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience & Biobehavioral Reviews*, *30*(2), 239–271. https://doi.org/10.1016/j.neubiorev.2005.07.001

Ellis, A. (1962). *Reason and emotion in psychotherapy.*

Endsley, M. R. (2012). Situation awareness. *Handbook of Human Factors and Ergonomics*, 553–568.

Frömer, R., Lin, H., Dean Wolf, C. K., Inzlicht, M., & Shenhav, A. (2021). Expectations of reward and efficacy guide cognitive control allocation. *Nature Communications*, *12*(1), 1030. https://doi.org/10.1038/s41467-021-21315-z

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Subirats Ferrer, M., Fischlmayr, I. C., ... Yamaguchi, S. (2011). Differences Between Tight and Loose Cultures: A 33-Nation Study. *Science*, *332*(6033), 1100–1104. https://doi.org/10.1126/science.1197754

Gibson, J. J. (2014). The theory of affordances:(1979). In *The people, place, and space reader* (pp. 56–60). Routledge.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(Volume 62, 2011), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. In *Advances in Experimental Social Psychology* (Vol. 38, pp. 69–119). Academic Press. https://doi.org/10.1016/S0065-2601(06)38002-1

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience, Vol 34*, *31*(1), 359–387.

Greene, J. (2014a). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.

Greene, J. D. (2014b). Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics. *Ethics*, *124*(4), 695–726. https://doi.org/10.1086/675875

Grossman, D. (2014). *On killing: The psychological cost of learning to kill in war and society*. Open Road Media.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Hansen, P. G. (2016). The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove? *European Journal of Risk Regulation*, *7*(1), 155–174. https://doi.org/10.1017/S1867299X00005468

Harmon-Jones, E., & Mills, J. (2019). An Introduction to Cognitive Dissonance Theory and an Overview of Current Perspectives on the Theory. In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology (2nd ed.).* American Psychological Association. https://doi.org/10.1037/0000135-000

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–100.

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, *36*, 427–440.

Johnson, E. J., Shu, S. B., Dellaert, B. G. C., Fox, C., Goldstein, D. G., Häubl, G., Larrick, R. P., Payne, J. W., Peters, E., Schkade, D., Wansink, B., & Weber, E. U. (2012). Beyond nudges: Tools of a choice architecture. *Marketing Letters*, *23*(2), 487–504. https://doi.org/10.1007/s11002-012-9186-1

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kandel, E., Schwartz, J., Jessell, T., Siegelbaum, S., & Hudspeth, A. J. (2013). *Principles of Neural Science* (5th ed.). McGraw-Hill Professional.

Karssing, E. (2023). The E-Word (Emotions) in Military Ethics Education: Making Use of the Dual-Process Model of Moral Psychology. In E.-H. Kramer & T. Molendijk (Eds.), *Violence in Extreme Conditions* (pp. 131–143). Springer International Publishing. https://doi.org/10.1007/978-3-031-16119-3_10

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology*, *31*(1), 457–501.

Kem, J. D. (2006). Use of the "Ethical Triangle" in Military Ethical Decision Making. *PAM*, *11*.

Kidder, R. M. (1995). *How good people make tough choices*. Morrow New York.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In *Decision making in action: Models and methods* (pp. 138–147). Ablex Publishing.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data–frame theory of sensemaking. In *Expertise out of context* (pp. 118–160). Psychology Press.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, *50*(5), 372–385. https://doi.org/10.1037/0003-066X.50.5.372

Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford University Press.

LeDoux, J. (2012). Rethinking the Emotional Brain. *Neuron*, *73*(4), 653–676. https://doi.org/10.1016/j.neuron.2012.02.004

Lessons and Doctrine Directorate ADF. (2021). *ADF Leadership* (3rd ed.). Lessons and

Doctrine Directorate.

Lin, H. (2024). Bayesian Epistemology. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2024). Metaphysics Research Lab, Stanford University.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1–25. https://doi.org/10.1086/392759

Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J. L., & Wagenmakers, E.-J. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences*, *119*(31), e2200300119. https://doi.org/10.1073/pnas.2200300119

Mandler, G. (1980). The Generation of Emotion: A Psychological Theory. In *Theories of Emotion* (pp. 219–243). Elsevier. https://doi.org/10.1016/B978-0-12-558701-3.50015-6

Mercier, H., & Sperber, D. (2017). *The Enigma of Reason*. Harvard University Press.

Messervey, D. L., Peach, J. M., Dean, W. H., & Nelson, E. A. (2023). Training for heat-of-the-moment thinking: Ethics training to prepare for operations. *Armed Forces & Society*, *49*(3), 593–611. https://doi.org/10.1177/0095327X221088325

Messervey, D., Dean, W. H., & Nelson, E. (2021). Making Moral Decisions Under Stress: A Revised Model for Defence. *Canadian Military Journal*, *21*(2).

Milgram, S. (1965). *Obedience [motion picture]*. Available from Penn State Media Sales.

Milgram, S., & Gudehus, C. (1974). *Obedience to authority*. Harper New York.

Minsky, M. (1986). *Society of mind*. Simon and Schuster.

Montoya, R. M., Horton, R. S., Vevea, J. L., Citkowicz, M., & Lauber, E. A. (2017). A re-examination of the mere exposure effect: The influence of repeated exposure on recognition, familiarity, and liking. *Psychological Bulletin*, *143*(5), 459–498. https://doi.org/10.1037/bul0000085

Moreno, J. L. (1934). *Who shall survive? A new approach to the problem of human interrelations*.

Musschenga, A. W. (2008). Moral Judgement and Moral Reasoning. In M. Düwell, C. Rehmann-Sutter, & D. Mieth (Eds.), *The Contingent Nature of Life: Bioethics and*

Limits of Human Existence (pp. 131–146). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6764-8_12

Oeberst, A., & Imhoff, R. (2023). Toward parsimony in bias research: A proposed common framework of belief-consistent information processing for a set of biases. *Perspectives on Psychological Science*, *18*(6), 1464–1487. https://doi.org/10.1177/17456916221148147

Opotow, S. (1990). *Moral exclusion and injustice: An introduction* (1; Vol. 46, pp. 1–20). Wiley Online Library.

Ordiway, B. (2022). Developing SOF Moral Reasoning. *Special Warfare*, *34*(3).

Osman, M., McLachlan, S., Fenton, N., Neil, M., Löfstedt, R., & Meder, B. (2020). Learning from Behavioural Changes That Fail. *Trends in Cognitive Sciences*, *24*(12), 969–980. https://doi.org/10.1016/j.tics.2020.09.009

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80. https://doi.org/10.1016/j.concog.2004.10.004

Patulny, R. V., & Lind Haase Svendsen, G. (2007). Exploring the social capital grid: Bonding, bridging, qualitative, quantitative. *International Journal of Sociology and Social Policy*, *27*(1/2), 32–51. https://doi.org/10.1108/01443330710722742

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*, *24*, 1–24.

Putnam, R. D. (1995). Bowling alone: America's declining social capital. In *The city reader* (pp. 188–196). Routledge.

Reicher, S. D., & Haslam, S. A. (2012). Change we can believe in: The role of social identity, cognitive alternatives, and leadership in group mobilization and social transformation. In *Culture and social change: Transforming society through the power of ideas* (pp. 53–73). IAP Information Age Publishing.

Reicher, S., & Haslam, S. A. (2013). Towards a 'Science of Movement': Identity, Authority and Influence in the Production of Social Stability and Social Change. *Journal of Social and Political Psychology*, *1*(1), 112–131. https://doi.org/10.5964/jspp.v1i1.266

Reicher, S., Haslam, S. A., & Hopkins, N. (2005). Social identity and the dynamics of leadership: Leaders and followers as collaborative agents in the transformation of social reality. *The Leadership Quarterly*, *16*(4), 547–568. https://doi.org/10.1016/j.leaqua.2005.06.007

Ross, L., & Nisbett, R. E. (2011). *The person and the situation: Perspectives of social psychology*. Pinter & Martin Publishers.

Ross, W. D. (2007). *The right and the good* (P. Stratton-Lake, Ed.; Reprinted). Clarendon Press.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(3), 273–294. https://doi.org/10.1016/0092-6566(77)90037-x

Sauer, H. (2019). *Moral thinking, fast and slow*. Routledge.

Schwabe, L., & Wolf, O. T. (2011). Stress-induced modulation of instrumental behavior: From goal-directed to habitual control of action. *Behavioural Brain Research*, *219*(2), 321–328.

Seeley, W. W. (2019). The salience network: A neural system for perceiving and responding to homeostatic demands. *Journal of Neuroscience*, *39*(50), 9878–9882.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews*, *68*, 651–668.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory. *In Two Minds: Dual Processes and Beyond*, 55–88.

Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1228–1248.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations.*

Brooks/Cole.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Thaler, R., & Sunstein, C. (2008). Nudge: Improving decisions about health, wealth and happiness. *Amsterdam Law Forum; HeinOnline: Online*, 89.

Thornton, P. H., & Ocasio, W. (2008). Institutional logics. *The Sage Handbook of Organizational Institutionalism*, *840*(2008), 99–128.

Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science*, *69*(2), 212–233.

Upton, C. L. (2009). Virtue Ethics and Moral Psychology: The Situationism Debate. *The Journal of Ethics*, *13*(2), 103–115. https://doi.org/10.1007/s10892-009-9054-2

Vincent, D. (2022). The Application of Ethical Leadership - The Vincent S-CALM Model. *Sandhurst Occasional Paper*, *36*.

Wilson, R. A., & Keil, F. (1998). The shadows and shallows of explanation. *Minds and Machines*, *8*, 137–159.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford university press.

Zillmann, D. (2008). Excitation Transfer Theory. In *The International Encyclopedia of Communication* (1st ed.). Wiley. https://doi.org/10.1002/9781405186407.wbiece049