

Self-image Interruption

Dr Zlatko Baracskai

1 Two readings of the title

The phrase admits two grammars, and the essay depends on holding both.

Reading A: action interrupts the self-image. Under sufficient task demand the self-narrative stops. Performers report it as a disappearance rather than a suppression: nothing was pushed away, the thing simply was not there for a period, and its absence was noticed only retrospectively. This is the condition variously named flow, absorption, being in the zone, or, in the vernacular of people who work with their hands and their attention at the same time, the tunnel. Its defining feature is not pleasure and not skill but the temporary unavailability of the self as an object of thought.

Reading B: the self-image interrupts action. The same relation, read backwards. A skilled sequence that runs correctly when unattended degrades when its owner begins to monitor it as a performance by a person of a certain description. The interruption is not distraction in the ordinary sense; the intruding content is specifically about the actor, and its arrival converts a single-task situation into a dual-task one.

The two readings are one mechanism, observed from either end. A finite processing system cannot simultaneously maintain a high-resolution model of the environment at the environment's tempo and a running evaluative model of itself. Whichever of the two holds the resource, the other is interrupted. Reading A describes the case where the world wins; Reading B describes the case where the biography does.

The essay's argument is that Reading B is the ordinary condition of adults in conditions of abundance, that it is largely useless as a control strategy and frequently harmful, and that Reading A is not a mystical state but an ordinary consequence of arranging one's environment so that the self-model loses the competition. The practical sections at the end describe how to detect the intrusion and how to audit choices for the pattern.

1.1 What is meant by “self-image”

A distinction is necessary, because much of what is loosely called “the self” is neither optional nor harmful.

Ulric Neisser (1928–2012) supplied the most usable partition. Neisser is the founding figure of the discipline: his textbook *Cognitive Psychology* (1967), written at Cornell, named the field and gave it its information-processing charter. His later career is unusual in that he spent it attacking his own creation. In *Cognition and Reality* (1976) he argued that laboratory cognitive psychology had purchased experimental control at the price of ecological validity, and he spent the following decades on real-world memory, most famously in the study with Nicole Harsch of flashbulb memories of the *Challenger* explosion, which demonstrated that vivid, confidently held autobiographical memories are frequently and grossly wrong. He also chaired the American Psychological Association task force that produced *Intelligence: Knowns and Unknowns* (1996), the field's most careful public statement on a subject that rewards carelessness. In “Five Kinds of Self-Knowledge” (*Philosophical Psychology*, 1988) he distinguished the **ecological self** (the self specified directly by optical, acoustic and proprioceptive information, present in the first

months of life, continuously updated, and never in doubt during action), the **interpersonal self** (specified by the reciprocal timing of exchange with another person), the **extended self** (the self of remembered past and anticipated future), the **private self** (the self of experiences known to be unshared), and the **conceptual self**: the theory one holds about oneself, including social roles, traits, values and explanations.

This essay's target is the conceptual and extended self in their evaluative mode. The ecological self is not a candidate for removal; it is the perceptual system's report on the perceiver's position in the layout, and it is the thing that makes action possible. When the tunnel closes, the ecological self is if anything sharpened. What goes quiet is the theory.

Call that theory the **self-image**: a low-bandwidth, verbally coded, evaluatively loaded, temporally extended summary of the organism, maintained for purposes of self-explanation and social presentation, and continuously updated by comparison against remembered standards and imagined audiences.

1.2 A note on evidential standards

The literature drawn on here has been through a decade of replication auditing, and several of the effects that a 2005-vintage essay would have cited without hesitation are now known to be smaller than advertised, contingent on unreported moderators, or absent. Where a claim rests on a contested finding, the contest is stated in the text rather than in a footnote. Where a mechanism survives its headline effect size, that is stated too. An essay arguing that self-flattering narratives are epistemically corrosive is obliged not to run on one.

2 The self-image is a model, and a poor one

Before the harms, the ontology. The self-image is not an organ, a homunculus, or a control centre. It is a representation, produced by machinery that produces representations, and it has the characteristic defects of a summary.

2.1 It is constructed after the fact

Michael Gazzaniga (b. 1939), now at the University of California, Santa Barbara, where he directs the SAGE Center for the Study of Mind, is the co-namer of cognitive neuroscience (with George A. Miller) and the principal investigator of the split-brain research programme begun under Roger Sperry at Caltech in the 1960s. His durable contribution to the present question is the **interpreter**: in patients whose cerebral commissures have been surgically divided, a stimulus presented to the right hemisphere can elicit a behaviour whose cause is unavailable to the speaking left hemisphere, and the left hemisphere does not report ignorance. It produces a fluent, confident, false explanation, and believes it. Gazzaniga and Joseph LeDoux set this out in *The Integrated Mind* (1978). The finding generalises past the surgical population: the narrating subsystem's function is to produce a coherent account, not an accurate one, and it has no access to a signal that would tell it which it has produced.

Richard Nisbett (b. 1941) and Timothy Wilson (b. 1951) established the same point in intact subjects. Nisbett, at the University of Michigan for his entire career, is known additionally for the fundamental work on lay statistical reasoning, for *The Geography of Thought* (2003) on cultural variation in cognition, and for the school-of-thought analysis of honour cultures. Wilson, at the University of Virginia, is the co-architect with Daniel Gilbert of the affective forecasting programme and the author of *Strangers to Ourselves* (2002). Their joint paper "Telling More Than We Can Know: Verbal Reports on Mental Processes" (*Psychological Review*, 1977) assembled the evidence that subjects report on their own cognitive processes by applying plausible causal

theories, not by introspecting the processes, and that these reports are accurate only when the plausible theory happens to be correct.

Petter Johansson and Lars Hall, at Lund University Cognitive Science in Sweden, sharpened this into an experimental paradigm. In “Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task” (*Science*, 2005) they had participants choose the more attractive of two faces, then, by sleight of hand, returned the *rejected* face and asked why it had been chosen. The manipulation went undetected on the large majority of trials, and participants produced detailed, confident justifications for a preference they had just declined; the phenomenon was named **choice blindness**. The team subsequently reproduced it for the taste of jam, the smell of tea, and, in 2012, for political attitudes, where participants defended reversed positions on policy questions minutes after endorsing the opposite.

Daryl Bem’s self-perception theory (*Advances in Experimental Social Psychology*, 1972), developed at Carnegie Mellon and Cornell, had already proposed the general form: when internal cues are weak or ambiguous, a person infers their own attitudes by observing their own behaviour and the circumstances in which it occurs, in the same way an outside observer would. The self-image is thus not the source of behaviour it is usually taken for. It is downstream of behaviour, and it reads the same public record everyone else does, with worse instrumentation and a stronger motive to distort.

2.2 It is a manipulable representation of a body

Thomas Metzinger (b. 1958), at Johannes Gutenberg University Mainz and founder of the MIND Group, argued in *Being No One* (2003) that the self is a **phenomenally transparent self-model**: transparent in the technical sense that the system cannot perceive it *as* a model, and therefore experiences its contents directly as reality. Transparency is what makes the self-image dangerous. A model recognised as a model is available for revision; a transparent one is simply the world.

The empirical demonstration came from Olaf Blanke’s laboratory at the École Polytechnique Fédérale de Lausanne. Blanke’s clinical work on out-of-body experiences produced by focal stimulation of the temporo-parietal junction, and then the “Video Ergo Sum” experiment (Lenggenhager, Tadi, Metzinger and Blanke, *Science*, 2007), in which stroking a participant’s back while they watched a video of a body being stroked in synchrony induced self-location drift toward the virtual body. The precursor is Botvinick and Cohen’s rubber hand illusion (*Nature*, 1998) and its extension by H. Henrik Ehrsson at the Karolinska Institutet to whole-body ownership and body-swapping. The bodily self, which feels least negotiable, is reconstructed continuously from multisensory correlations and can be moved by a few seconds of contrived stimulation.

If the self-model’s most concrete stratum is that pliable, the conceptual stratum — assembled from remembered episodes that are themselves reconstructive, and from social feedback that is noisy and strategic — has no claim to be treated as data.

2.3 The bandwidth argument

Whatever else it is, the self-image is verbal, sequential and slow. Perception and action are parallel and fast. Harold Pashler, at the University of California, San Diego, established the strongest version of the structural constraint. Pashler is also, not incidentally, one of the earliest and most effective critics of the field’s statistical practice, and a founder of the reform movement that produced the replication audits cited throughout this essay. His review “Dual-Task Interference in Simple Tasks: Data and Theory” (*Psychological Bulletin*, 1994) consolidated the evidence for a **central bottleneck**: when two tasks each require a decision stage, those stages cannot overlap, and the second is postponed. Stephen Monsell, at Exeter and previously the Medical

Research Council’s Applied Psychology Unit in Cambridge, documented the associated **switch cost** (“Task Switching”, *Trends in Cognitive Sciences*, 2003): reconfiguring the system between task sets costs measurable time even when the switch is fully anticipated and the interval to prepare is long. There is a residual cost that preparation does not abolish.

The consequence for the present argument is not metaphorical. Self-referential thought is a task. Entering it and leaving it incurs the standard costs. Any activity whose demands arrive faster than the switch cost cannot be performed while it is running.

3 The uselessness argument: the self-image does not do the work claimed for it

The folk theory holds that a favourable, stable self-image is a precondition of competent action; that people achieve because they believe in themselves; and that the maintenance of this belief is therefore a legitimate first call on attention. Each clause is either false or backwards.

3.1 Self-esteem is an output, not an input

The definitive audit is Roy Baumeister, Jennifer Campbell, Joachim Krueger and Kathleen Vohs, “Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles?” (*Psychological Science in the Public Interest*, 2003). Baumeister (b. 1953) trained at Princeton under Edward E. Jones and held chairs at Case Western Reserve, Florida State and the University of Queensland; his wider contributions include the need-to-belong framework (with Mark Leary, 1995), the analysis of self-regulation as a limited resource, and the bad-is-stronger-than-good asymmetry. The self-esteem review was commissioned by an APA-affiliated journal specifically to settle a question on which public policy had already been made. Its findings: the correlation between self-esteem and academic achievement is real but modest, and longitudinal analyses indicate that achievement causes self-esteem rather than the reverse; self-esteem does not predict occupational success once ability is controlled; interventions that raise self-esteem without raising competence do not raise performance and in some samples lower it; the reliable correlates of high self-esteem are happiness and initiative, both plausibly consequences.

The review is worth taking on its own terms and also worth reading in the light of the subsequent replication crisis, which severely damaged Baumeister’s other major programme, ego depletion (a 23-laboratory registered replication led by Martin Hagger, *Perspectives on Psychological Science*, 2016, found an effect size indistinguishable from zero). The self-esteem review, being a synthesis of correlational and longitudinal literatures rather than a set of small-N laboratory demonstrations, has aged considerably better than the resource model. Its conclusion has not been overturned: the self-image is a gauge that has been mistaken for an engine.

3.2 Gauges make bad targets

Mark Leary (b. 1954), at Duke University after a long tenure at Wake Forest, supplied the functional account. Leary’s wider work includes the need-to-belong theory with Baumeister, the social anxiety literature, and the early psychometric work on self-compassion with Kristin Neff. In Leary, Tambor, Terdal and Downs, “Self-Esteem as an Interpersonal Monitor: The Sociometer Hypothesis” (*Journal of Personality and Social Psychology*, 1995), self-esteem is analysed as the readout of a system that monitors one’s relational value to others: it tracks inclusion, and its function is to make impending exclusion aversive enough to prompt corrective action. On this account self-esteem is a fuel gauge. Leary’s later book *The Curse of the Self* (2004) draws the obvious inference, which is that the same self-reflective capacity that permits planning and

moral reasoning is also the substrate of chronic suffering, and that its costs are systematically underestimated because it is the faculty doing the estimating.

A gauge that becomes a target ceases to measure. Pursuing high self-esteem directly optimises the readout — through selective attention to favourable evidence, avoidance of diagnostic tests, self-handicapping, and the recruitment of audiences chosen for their compliance — while degrading the quantity it was built to track.

Jennifer Crocker made the cost explicit. Crocker, at the University of Michigan and subsequently Ohio State, developed **contingencies of self-worth** with Connie Wolfe (*Psychological Review*, 2001): self-esteem is not a global quantity but a set of domain-specific bets, and what matters is which domains a person has staked. In Crocker and Lora Park, “The Costly Pursuit of Self-Esteem” (*Psychological Bulletin*, 2004), the documented costs of pursuing self-worth in a staked domain include impaired learning (because failure must be avoided rather than analysed), impaired relationships (because others become instruments of validation), impaired autonomy and self-regulation, and elevated stress reactivity. The costs scale with how much of the self-image rides on the domain.

Michael Kernis (1953–2009), at the University of Georgia, added the stability dimension. In a programme summarised in “Toward a Conceptualization of Optimal Self-Esteem” (*Psychological Inquiry*, 2003), Kernis showed that self-esteem *level* is a poor predictor once *fragility* is measured. Fragile high self-esteem — high in mean, unstable day to day, contingent on outcomes, and defended — predicts hostility, defensiveness and anger in response to threat more strongly than low self-esteem does. Secure high self-esteem, which does not require defence, predicts none of it. What is pathological is not a low opinion but a load-bearing one.

3.3 The best available evidence that the narrative is not the active ingredient

The strongest single result for this essay’s thesis comes from clinical trials, not from social psychology.

Neil Jacobson (1949–1999) at the University of Washington was a marital-therapy researcher who turned to depression and to the unfashionable question of which components of an effective treatment are actually doing the work. Cognitive therapy for depression, as developed by Aaron Beck, has three nested components: behavioural activation (scheduling contact with reinforcing activity), modification of automatic thoughts, and modification of core beliefs about the self. Jacobson and colleagues dismantled it. In “A Component Analysis of Cognitive-Behavioral Treatment for Depression” (*Journal of Consulting and Clinical Psychology*, 1996), 150 depressed patients were randomised to the full package, to activation plus automatic-thought work, or to **behavioural activation alone**, with no work on the self-concept whatsoever. The three conditions were equivalent at termination and at six-month follow-up. The component that addressed the self-narrative was not necessary for the outcome.

The finding was not a fluke of one laboratory. Sona Dimidjian, now at the University of Colorado Boulder, led the placebo-controlled replication with Steven Hollon and Keith Dobson (*JCCP*, 2006), in which behavioural activation matched antidepressant medication among the more severely depressed and outperformed cognitive therapy in that stratum. The COBRA trial (Richards et al., *The Lancet*, 2016), a large UK non-inferiority trial, found behavioural activation delivered by junior mental-health workers non-inferior to full cognitive-behavioural therapy delivered by qualified therapists, at roughly a fifth less cost.

Read plainly: the reliable route out of the most common disorder of self-narrative is to change what the person does in their environment and to leave the narrative alone. It reorganises itself, or it does not and stops mattering.

4 Harm, first form: competent minds falling prey to inhibition

The clearest evidence that the self-image is not merely useless but subtractive comes from cases where a demonstrably functional system is degraded by the arrival of self-referential processing, and by nothing else. The manipulations are small, the populations are unimpaired, and the losses are large.

4.1 Choking: the experimental case

Roy Baumeister’s “Choking Under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance” (*JPSP*, 1984) established the paradigm and named the mechanism. Baumeister’s claim was that pressure induces attention to the self as performer, and that this attention is what disrupts execution — not arousal, not anxiety, and not the incentive as such. In his experiments, incentives large enough to matter reliably impaired performance on a skilled perceptual-motor task, and the impairment tracked dispositional and situationally induced self-consciousness.

Sian Beilock supplied the decisive dissociation. Beilock (b. 1976) took her doctorate at Michigan State, established the Human Performance Laboratory at the University of Chicago, and subsequently became president of Barnard College and then of Dartmouth College; her wider research concerns embodied mathematics learning, the transmission of maths anxiety from teachers to pupils, and the role of sensorimotor experience in comprehension. Her book *Choke* (2010) is the field’s summary. Two papers carry the argument.

In Beilock and Thomas Carr, “On the Fragility of Skilled Performance: What Governs Choking Under Pressure?” (*Journal of Experimental Psychology: General*, 2001), golf putting practised to automaticity was disrupted by pressure, and the disruption was specific: skill-focused attention, not distraction, was the operative variable. In Beilock, Carr, MacMahon and Starkes, “When Paying Attention Becomes Counterproductive” (*JEP: Applied*, 2002), the crossover was demonstrated directly. Novices and experts performed golf putting and soccer dribbling under two secondary loads: a **dual task** (monitoring an auditory stream, which consumes attention but directs none of it at the skill) and a **skill focus** (attending to a component of one’s own movement). Novices were impaired by the dual task and helped, or unaffected, by skill focus. Experts showed the reverse: they were unimpaired by the dual task and significantly impaired by attending to their own execution.

This is the structural result on which the essay’s first harm rests. The same instruction that assists a beginner damages an expert, because expertise consists in having compiled the sequence into a form that does not require, and cannot tolerate, step-by-step supervision. Self-attention does not add a resource. It reinstates a control regime the system has outgrown.

4.2 The attentional-focus literature

Gabriele Wulf, at the University of Nevada, Las Vegas, has spent three decades on a variable with an unreasonable effect size for so small a manipulation. Wulf trained in Germany under Wolfgang Prinz and worked at the Max Planck Institute for Psychological Research in Munich before moving to the United States; her legacy is the systematic demonstration that **where attention is directed** dominates almost everything else in motor learning instruction, and, with Rebecca Lewthwaite, the OPTIMAL theory of motor learning (*Psychonomic Bulletin & Review*, 2016), which integrates attentional focus with expectancies and autonomy support.

The founding study is Wulf, Höß and Prinz, “Instructions for Motor Learning: Differential Effects of Internal Versus Focus of Attention” (*Journal of Motor Behavior*, 1998), using a ski

simulator and a stabilometer. An **internal focus** directs attention to the body's own movement ("push with your feet"); an **external focus** directs it to the movement's effect in the environment ("push on the markers"). External focus produced faster acquisition, better retention, and better transfer, at every skill level tested. Wulf's 2013 review in the *International Review of Sport and Exercise Psychology* assembled the subsequent fifteen years: the advantage appears in accuracy, in force production, in movement economy measured by oxygen consumption and electromyography, in balance, and in patients with Parkinson's disease and stroke.

The interpretation Wulf and Prinz favour is the **constrained action hypothesis**: an internal focus induces conscious intervention in processes that self-organise more efficiently without it, and the intervention constrains the motor system. Prinz's own theoretical contribution, **common coding** (*European Journal of Cognitive Psychology*, 1997), makes this coherent: perception and action are represented in a shared code in terms of their distal effects, so an instruction phrased in terms of environmental effects addresses the motor system in its native format, and one phrased in terms of body parts does not.

Note what an internal focus *is*, in this essay's terms. It is the smallest possible self-image: attention to one's own body as an object of evaluation. It costs measurable oxygen.

4.3 Verbalisation as degradation

Jonathan Schooler, at the University of California, Santa Barbara, is known for two programmes, both relevant. The first is **verbal overshadowing**, established with Tonya Engstler-Schooler (*Cognitive Psychology*, 1990): describing a previously seen face in words impairs subsequent recognition of that face relative to an unrelated verbal task. The demonstration mattered because it showed that the act of rendering a perceptual representation into a propositional one can overwrite or interfere with the perceptual one. It also mattered later for a second reason: verbal overshadowing was chosen as the subject of the first **Registered Replication Report** (Alogna et al., *Perspectives on Psychological Science*, 2014), a coordinated 31-laboratory effort. The effect replicated, at roughly a third of the original magnitude. This is the honest shape of the finding: real, smaller than advertised, and directionally as Schooler described.

Schooler's second programme, with Jonathan Smallwood (now at Queen's University, Kingston, formerly at the University of York), is the science of mind-wandering. Their *Annual Review of Psychology* synthesis (2015) and the earlier "Meta-Awareness, Perceptual Decoupling and the Wandering Mind" (*Trends in Cognitive Sciences*, 2011) introduced the term that this essay needs most: **perceptual decoupling**. When attention shifts to internally generated content, processing of sensory input is attenuated — measurably, in reduced amplitude of stimulus-evoked cortical potentials and in degraded performance on the ostensible task — and, critically, the subject is typically unaware that it has happened. Meta-awareness of mind-wandering lags the wandering itself. The interruption is silent.

4.4 The mind-wandering cost, measured in the field

Matthew Killingsworth and Daniel Gilbert quantified the affective cost. Gilbert, at Harvard, is the co-founder with Timothy Wilson of the affective forecasting literature — the systematic demonstration that people mispredict the intensity and duration of their own future emotional responses, documented in *Stumbling on Happiness* (2006) and grounded in the immune-neglect and impact-bias findings of the late 1990s. Killingsworth, now at the Wharton School, built an experience-sampling application that prompted 2,250 adults at random moments to report what they were doing, whether they were thinking about something other than their current activity, and how they felt. "A Wandering Mind Is an Unhappy Mind" (*Science*, 2010) reported that minds wandered on 46.9% of samples, that this was true of every activity except sex, that wandering predicted lower subsequent happiness, and — using the time-lagged structure — that

the direction of the relation ran from wandering to unhappiness rather than the reverse. Notably, the content of the wandering mattered less than the fact of it: pleasant off-task thought was no better than being on task.

4.5 Self-focus in the clinic: social anxiety

The most complete clinical model of self-image interruption belongs to David M. Clark and Adrian Wells. Clark, Professor of Experimental Psychology at Oxford and previously at the Institute of Psychiatry in London, is the author of the cognitive models of panic disorder (1986) and social phobia, and, with the economist Richard Layard, the principal architect of the English Improving Access to Psychological Therapies programme, the largest routine-outcome-monitored psychological treatment service in the world. Wells, at the University of Manchester, is the originator of **metacognitive therapy** and, with Gerald Matthews, of the Self-Regulatory Executive Function (S-REF) model (*Attention and Emotion*, 1994).

Their joint chapter, “A Cognitive Model of Social Phobia” (in Heimberg et al., eds., *Social Phobia*, 1995), proposes that on entering a feared social situation the person shifts attention inward and constructs an image of how they appear to others, then **treats that image as evidence about how they actually appear**. Two consequences follow. First, the attentional resources needed to monitor the actual social situation are consumed by the internal model, so the person acquires no disconfirming data and performs worse. Second, the model is generated from anxiety-related interoceptive cues, so it is systematically worse than the reality.

Hackmann, Clark and McManus (“Recurrent Images and Early Memories in Social Phobia”, *Behaviour Research and Therapy*, 2000) documented the phenomenology: socially anxious patients report recurrent images of themselves in social situations, and these images are predominantly taken from the **observer perspective** — the patient sees themselves from outside, as a third party would. The perspective is diagnostic. A person absorbed in a situation sees the situation; a person absorbed in a self-image sees themselves.

Wells generalised the pattern beyond social anxiety into the **cognitive attentional syndrome** (CAS): the transdiagnostic combination of perseverative thinking (worry and rumination), threat monitoring, and coping strategies that prevent disconfirmation. In the metacognitive account, the content of the negative self-belief is irrelevant to treatment; what maintains disorder is the *style* of attending, and the therapeutic target is therefore the process, not the proposition. His Attention Training Technique trains externally directed, flexibly divided auditory attention explicitly in order to interrupt self-focus. The evidence base for metacognitive therapy is smaller than that for standard cognitive-behavioural therapy and largely from the originating group and its collaborators, which is a reason for calibration rather than dismissal; the descriptive model of self-focused attention is much more securely established than the specific treatment package built on it.

4.6 Stereotype threat, stated carefully

Claude Steele (b. 1946) held chairs at Michigan, Stanford, Columbia and Berkeley, serving as provost at the latter two; his wider contributions include **self-affirmation theory** (1988), which showed that threats to self-integrity in one domain can be neutralised by affirming values in another, and the alcohol myopia model of intoxicated cognition. With Joshua Aronson (now at New York University) he published “Stereotype Threat and the Intellectual Test Performance of African Americans” (*JPSP*, 1995): describing a difficult verbal test as diagnostic of ability depressed the performance of Black undergraduates relative to a non-diagnostic framing, with no corresponding effect for white undergraduates. The proposed mechanism is that a self-relevant, socially charged concern occupies working memory during the task.

The honest position on this literature in the 2020s is as follows. The mechanism has good independent support: Toni Schmader (University of British Columbia) and Michael Johns, “Converging Evidence That Stereotype Threat Reduces Working Memory Capacity” (*JPSP*, 2003), showed the mediating capacity reduction directly. The headline effect, however, has not survived meta-analytic scrutiny intact. Flore and Wicherts (*Journal of School Psychology*, 2015) found substantial evidence of publication bias in the schoolgirl-mathematics literature and a corrected effect near zero; Shewach, Sackett and Quint (*Journal of Applied Psychology*, 2019) found the effect to be small and largely confined to laboratory conditions with weak analogues in operational testing. The correct inference is not that the phenomenon is fictional but that its field magnitude is modest and heterogeneous. For the present argument, what matters is the mechanism rather than the size of any one demonstration: a concern about what one’s performance will mean about the kind of person one is competes for the same capacity the performance requires. Stereotype threat is a socially inflicted instance of a general form.

4.7 Rumination

Susan Nolen-Hoeksema (1959–2013), at Yale after positions at Stanford and Michigan, is the author of **response styles theory** and the single most influential researcher on rumination. Her wider contributions include the demonstration that the gender difference in depression prevalence is substantially mediated by ruminative response style, and the extension of rumination to eating disorders and substance use. “Rethinking Rumination” (with Wisco and Lyubomirsky, *Perspectives on Psychological Science*, 2008) is the definitive review: rumination is repetitive, passive, abstract thought focused on the causes and consequences of one’s distress and on one’s own inadequacy; it prospectively predicts onset, severity and duration of depressive episodes; and, contrary to the folk belief that it constitutes problem-solving, it degrades problem-solving, impairs instrumental behaviour, and drives away social support.

Rumination is the self-image in its purest procedural form. It is entirely about the self, it produces no action, it terminates in a verdict, it feels obligatory, and it makes its subject demonstrably worse at everything the situation requires.

4.8 The neural correlate, with a caveat

Marcus Raichle (b. 1937), at Washington University in St. Louis, is a founder of functional brain imaging: his laboratory developed much of the positron emission tomography methodology for measuring regional blood flow, and he named the **default mode network** in “A Default Mode of Brain Function” (*PNAS*, 2001) after noticing that a consistent set of midline and lateral parietal regions was reliably *deactivated* by externally directed tasks, and consumed a disproportionate share of the brain’s energy budget at rest. The DMN’s principal correlates are self-referential processing, autobiographical memory, mental time travel and social cognition.

Norman Farb and Zindel Segal, at the University of Toronto, made the essential dissociation. Segal, at Toronto Scarborough, is the co-developer with Mark Williams and John Teasdale of Mindfulness-Based Cognitive Therapy, whose relapse-prevention efficacy in recurrent depression is established across several trials and a patient-level meta-analysis. In Farb et al., “Attending to the Present: Mindfulness Meditation Reveals Distinct Neural Modes of Self-Reference” (*Social Cognitive and Affective Neuroscience*, 2007), participants performed a trait-adjective task under two instructional sets: **narrative focus**, evaluating whether traits described them, and **experiential focus**, attending to present-moment sensation without elaboration. In untrained participants the two conditions produced overlapping midline prefrontal activity — that is, the experiential instruction did not much succeed. In participants who had completed an eight-week mindfulness course, the two modes dissociated: experiential focus produced reduced midline

prefrontal activity and increased lateral prefrontal, insular and somatosensory activity, and the normally strong coupling between midline and insular regions was uncoupled.

The finding is important because it identifies the narrative self and the momentary experiential self as **separable** rather than as two descriptions of one thing, and because the separation is trainable. It should be read with the standard caveats: small samples by current standards, a between-groups comparison with self-selection into training, and the general fragility of reverse inference from regional activation.

The flow literature’s neural counterpart carries a similar caveat. Arne Dietrich, at the American University of Beirut, proposed the **transient hypofrontality hypothesis** (*Consciousness and Cognition*, 2003; extended to flow in 2004): that states of effortless absorption involve a downregulation of prefrontal functions, including the explicit self-model, because attentional resources are monopolised by the task’s sensorimotor demands. Dietrich’s wider legacy is largely deflationary — he is among the sharpest critics of the neuroscience of creativity and of the “mad genius” literature — and he applied the same scepticism here. The direct tests are equivocal. Martin Ulrich and colleagues at Ulm (“Neural Correlates of Experimentally Induced Flow Experiences”, *NeuroImage*, 2014), using difficulty-matched mental arithmetic, found *decreased* activity in medial prefrontal cortex and amygdala during flow alongside *increased* activity in inferior frontal gyrus and putamen. Global hypofrontality is not supported; selective deactivation of self-referential and threat-appraisal circuitry is. That narrower claim is exactly the one this essay needs, and it is the one the data support.

Charles Limb and Allen Braun provided the performance-domain demonstration. Limb, an otolaryngologist and jazz musician, then at Johns Hopkins and now at the University of California, San Francisco, where he directs the Douglas Grant Cochlear Implant Center; Braun then at the National Institute on Deafness and Other Communication Disorders. In “Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation” (*PLoS ONE*, 2008), professional pianists improvised on a plastic keyboard in the scanner. Improvisation relative to rehearsed performance produced **deactivation of dorsolateral prefrontal and lateral orbitofrontal cortex** — regions associated with self-monitoring and evaluation — alongside activation of medial prefrontal cortex, associated with self-generated, internally motivated action. The authors described the pattern as dissociated self-monitoring and self-expression. The sample was six musicians, and the reader should weight it accordingly; but the direction agrees with the behavioural literature that arrives at the same place from five other approaches.

5 Harm, second form: the established preference as a debility

The self-image is not only a monitor. It is also a store of settled preferences, and a preference is a compression: it substitutes a stored verdict for a fresh evaluation. That trade is favourable when the environment is stationary and unfavourable otherwise, and the self-image is unusually bad at noticing which case obtains, because the preference has been incorporated into the description of who one is and is therefore defended rather than tested.

5.1 Einstellung: the good solution blocks the better one

Abraham Luchins (1914–2005), a student of Max Wertheimer and later at the State University of New York at Albany, ran the canonical demonstration in “Mechanization in Problem Solving: The Effect of *Einstellung*” (*Psychological Monographs*, 1942). Participants solved a series of water-jar problems all soluble by one moderately complex formula, then received problems soluble both by that formula and by a trivially simpler one. The great majority used the complex method; a

substantial minority declared insoluble a problem that a naive control group solved immediately. The learned solution did not merely compete with the better one. It concealed it.

Merim Bilalić and Peter McLeod, then at Oxford, working with Fernand Gobet, showed the mechanism in experts using eye tracking. In “Why Good Thoughts Block Better Ones: The Mechanism of the Pernicious Einstellung (Set) Effect” (*Cognition*, 2008), chess players were given positions containing both a familiar smothered-mate solution and a shorter, faster win. Strong players who found the familiar solution overwhelmingly failed to find the better one — and their eye movements showed that they continued to fixate the squares relevant to the familiar solution while sincerely reporting that they were searching for alternatives. The block was not a failure of motivation or of honesty. It operated below the level at which the players had access. Bilalić, now at Northumbria University, has since extended the work to medical diagnosis, where the same structure produces premature closure.

This is the exact model of a self-image. A settled account of who one is directs attention to the evidence that account makes salient, and the search feels open while it is closed.

5.2 Functional fixedness, and its absence in children

Karl Duncker (1903–1940), a Gestalt psychologist driven out of Germany in 1935 who died in the United States at thirty-seven, described **functional fixedness** in *On Problem-Solving* (1945): an object’s established function inhibits its use in a novel one. In the candle problem, a box of tacks is perceived as a container and therefore not as a shelf.

The developmental extension is the striking part. Tim German, at the University of California, Santa Barbara, and Margaret Anne Defeyter, in “Immunity to Functional Fixedness in Young Children” (*Psychonomic Bulletin & Review*, 2000), gave a box-and-tacks problem to children of five, six and seven. **Five-year-olds were immune**: they solved the pre-utilisation and control versions equally fast. Six- and seven-year-olds showed the adult effect. The impairment is acquired. It arrives with the accumulation of stable function-knowledge, which is to say with competence.

The generalisation to the present subject is direct and unsentimental. A self-image is a stable function-assignment applied to oneself. Its acquisition is a developmental achievement and its consequence is a narrowing of the set of uses to which one can be put.

5.3 Habit outrunning motive

Wendy Wood, Provost Professor at the University of Southern California, is the principal contemporary theorist of habit; her framework with David Neal (*Psychological Review*, 2007) treats habits as direct context-response associations that bypass goal mediation once formed, and her book *Good Habits, Bad Habits* (2019) summarises the programme. The cleanest experimental illustration is Neal, Wood, Wu and Kurlander, “The Pull of the Past: When Do Habits Persist Despite Conflict with Motives?” (*Personality and Social Psychology Bulletin*, 2011). Cinema-goers were given popcorn that was either fresh or a week stale. Participants with weak popcorn-eating habits ate substantially less of the stale popcorn, as sense requires. Participants with strong habits ate as much stale as fresh, and rated it as unpleasant while eating it. Moved to a meeting room — the same people, the same popcorn, a different context — strong-habit participants discriminated normally.

The result isolates the variable. It is not appetite and not judgement; the context cue retrieves the response, and the response executes against the current evaluation. An established preference does not merely persist in the absence of evidence. It persists in the presence of contrary evidence that the person is concurrently reporting.

5.4 Cognitive entrenchment

Erik Dane, then at Rice University’s Jones Graduate School of Business, formalised the organisational version in “Reconsidering the Trade-Off Between Expertise and Flexibility: A Cognitive Entrenchment Perspective” (*Academy of Management Review*, 2010). Entrenchment is a high level of stability in one’s domain schemas. It is what makes expert performance fast and reliable within the domain, and it is simultaneously what degrades adaptive performance when the domain shifts and what impairs radical creativity and analogical transfer out of the domain. Dane’s contribution is to treat this as a trade-off with identifiable moderators — task variety, breadth of attention, deliberate exposure to alternative domains — rather than as an unavoidable tax on competence.

5.5 Self-verification: preferring a true bad account to a false good one

William Swann, at the University of Texas at Austin, established the most counterintuitive and, for this essay, the most damning result about the self-image. Swann’s wider contributions include the identity fusion construct and its application to extreme pro-group behaviour, and the book *Self-Traps* (1996). **Self-verification theory** holds that people are motivated to confirm their existing self-views, including negative ones, because a stable self-concept underwrites the predictability of social interaction.

In Swann, Pelham and Krull, “Agreeable Fancy or Disagreeable Truth? Reconciling Self-Enhancement and Self-Verification” (*JPSP*, 1989), participants chose which of two evaluators to interact with. People with negative self-views in a domain preferentially chose the evaluator whose appraisal was unfavourable — and reported that the unfavourable evaluator seemed more perceptive. In Swann, Hixon and De La Ronde (“Embracing the Bitter Truth”, *Psychological Science*, 1992), married people with negative self-views were **more committed** to spouses who appraised them unfavourably, and withdrew from spouses who appraised them well.

This is the crucial finding for the argument that the self-image is not a servant of well-being. Given a choice between accurate self-continuity and a better life, a substantial fraction of people choose continuity, and experience the choice as discernment.

5.6 The consistency engine

Leon Festinger (1919–1989), who worked at Minnesota, Stanford and the New School for Social Research and whose wider legacy includes social comparison theory (1954) and, in his last career, archaeology and the psychology of technological change, provided the motivational engine in *A Theory of Cognitive Dissonance* (1957). Festinger and Carlsmith’s forced-compliance experiment (*Journal of Abnormal and Social Psychology*, 1959) showed that participants paid one dollar to describe a tedious task as enjoyable subsequently rated it as more enjoyable than those paid twenty; insufficient external justification produces internal revision. Elliot Aronson’s later reformulation at Texas and the University of California, Santa Cruz specified what is actually dissonant: not any two inconsistent cognitions, but inconsistency between behaviour and the **self-concept** as competent and moral.

Aronson’s revision is the point. Dissonance is a self-image maintenance system. Its running cost is the systematic distortion of the evidential record whenever the record and the self-description conflict, and it resolves the conflict in whichever direction is cheaper, which is almost always the record.

5.7 A theoretical frame: precision on the wrong prior

Karl Friston, at University College London’s Wellcome Centre for Human Neuroimaging, is the author of statistical parametric mapping and dynamic causal modelling — the analytic infrastructure of most of the neuroimaging literature cited above — and of the free energy principle (“The Free-Energy Principle: A Unified Brain Theory?”, *Nature Reviews Neuroscience*, 2010). Andy Clark, at the University of Sussex after chairs at Washington University in St. Louis and Edinburgh, is the co-author with David Chalmers of “The Extended Mind” (*Analysis*, 1998) and the principal expositor of predictive processing for cognitive science (*Whatever Next?*, *Behavioral and Brain Sciences*, 2013; *Surfing Uncertainty*, 2016).

On the predictive-processing account, perception and action minimise prediction error, and the relative influence of prior models versus incoming evidence is set by **precision weighting**: an estimate of how reliable each is. The self-model, on this reading, is a set of high-level priors. Two failure modes follow directly. If precision on the self-model is set too high, sensory and social evidence cannot revise it, and the system instead acts to bring the world into conformity — which is a formally exact description of self-verification. If precision on self-related prediction errors is set too high, ordinary social noise is read as diagnostic information about the self, which is a formally exact description of the socially anxious self-monitor.

This is a framework, not a result, and the free energy principle’s empirical content is genuinely disputed. It is offered here because it renders the preceding six subsections as instances of one parameter rather than six unrelated phenomena, and because it makes the intervention specification legible: what is wanted is not a better self-model but a lower precision on the one there is.

6 The falling threshold: will to sustain discomfort as a decaying quantity

A self-image requires maintenance, and maintenance requires that threats to it be avoidable. The general form of the resulting behaviour is avoidance of aversive internal states, and avoidance has a formal property that makes it self-accelerating: it is negatively reinforced. The removal of an unpleasant state immediately follows the avoidant act, so the act is strengthened, and it is strengthened whether or not the avoided state was actually dangerous. There is no mechanism by which the organism learns that the avoidance was unnecessary, because the test is never run.

6.1 Experiential avoidance

Steven Hayes, at the University of Nevada, Reno, is the originator of Acceptance and Commitment Therapy and, with Dermot Barnes-Holmes and Bryan Roche, of Relational Frame Theory, the post-Skinnerian behavioural account of language and cognition. The relevant paper is Hayes, Wilson, Gifford, Follette and Strosahl, “Experiential Avoidance and Behavioral Disorders: A Functional Dimensional Approach to Diagnosis and Treatment” (*Journal of Consulting and Clinical Psychology*, 1996), which proposed **experiential avoidance** — unwillingness to remain in contact with particular private experiences, and action taken to alter their form or frequency — as a functional dimension cutting across the diagnostic categories. Its evidential status is now good: measures of experiential avoidance and psychological inflexibility predict a wide range of outcomes, and reduction in inflexibility mediates ACT’s treatment effects in a number of trials.

Hayes’s terminology is unusually well suited to this essay. ACT distinguishes **self-as-content** — the conceptualised self, the narrative one defends — from **self-as-context**, the perspective from which experience is observed. The clinical target is not to improve the content but to reduce **cognitive fusion**: the condition in which a thought is responded to as though it were the state

of affairs it describes. Fusion is Metzinger’s transparency restated in behavioural terms, and defusion is the operational form of the essay’s recommendation.

6.2 The threshold falls: opponent processes and allostasis

Richard Solomon (1918–2002), at the University of Pennsylvania, where he also supervised the learned-helplessness work of Seligman and Maier, published with John Corbit “An Opponent-Process Theory of Motivation” (*Psychological Review*, 1974). Any affective state, pleasant or aversive, elicits a slave opponent process of opposite sign. With repetition, the primary process weakens and the opponent process strengthens, is recruited faster, and decays more slowly. The observable consequence is that the pleasure of a repeated comfort shrinks while the discomfort of its absence grows.

George Koob, director of the National Institute on Alcohol Abuse and Alcoholism and previously at the Scripps Research Institute, and Michel Le Moal at the University of Bordeaux, gave the mechanism a neurobiological form in “Drug Abuse: Hedonic Homeostatic Dysregulation” (*Science*, 1997) and the subsequent **allostasis** model. Their argument is that repeated recruitment of reward systems produces compensatory recruitment of anti-reward systems (corticotropin-releasing factor, dynorphin), and the set point around which the system regulates drifts. The organism is not returning to baseline between episodes; the baseline is moving. Koob’s wider legacy is the extension of this framework from drugs to compulsive behaviours generally.

Applied to ordinary conduct rather than to addiction, the model predicts precisely the phenomenon named in this essay’s brief. The threshold of will required to sustain a given discomfort is not a fixed personal characteristic. It is a moving quantity that falls with each avoidance, because the avoided discomfort’s opponent has been strengthened and the ambient set point has shifted. A person who has not been bored in three years does not have the same boredom-tolerance they had three years ago, minus a bit of practice. They have a different set point, at which unremarkable understimulation registers as aversive.

6.3 Fifteen minutes alone

Timothy Wilson’s laboratory produced the demonstration that made the point unignorable. “Just Think: The Challenges of the Disengaged Mind” (Wilson, Reinhard, Westgate, Gilbert, Ellerbeck, Hahn, Brown and Shaked, *Science*, 2014) reported eleven studies in which participants were asked to sit alone in an unadorned room, without possessions or devices, and entertain themselves with their thoughts for six to fifteen minutes. They found it difficult and unenjoyable, whether in the laboratory or at home, across ages from 18 to 77. In the final study participants had first sampled a mild electric shock and had, in most cases, stated they would pay money to avoid receiving it again. Left alone with their thoughts for fifteen minutes with the shock apparatus available, **67% of men and 25% of women shocked themselves at least once**; one outlier aside, the modal self-shocker administered several.

The finding is often reported as evidence that people cannot bear their own company. The more precise reading is that unstructured time is the condition in which the self-image has no competitor, and that being alone with an evaluative self-model is aversive enough to be worth paying to interrupt. The participants were not fleeing silence. They were fleeing a narrator.

6.4 Boredom is an attentional failure, not a lack of stimulation

John Eastwood at York University in Toronto and James Danckert at the University of Waterloo have argued for a definition that matters here. In Eastwood, Frischen, Fenske and Smilek, “The Unengaged Mind: Defining Boredom in Terms of Attention” (*Perspectives on Psychological Science*, 2012), and in Danckert and Eastwood’s *Out of My Skull* (2020), boredom is characterised

as the aversive state of *wanting, but being unable, to engage in satisfying activity* — a failure of the attentional system to lock onto something, coupled with awareness of that failure. It is not caused by the environment being poor in stimulation; it is a mismatch between available engagement and the capacity to take it up, and it is best understood as a signal calling for agency rather than for input.

This matters because the standard response to boredom — supplying more stimulation — treats the state as a deficiency of input, which strengthens the association between the aversive state and immediate consumption, and does nothing to restore the capacity to engage. The alternative response, tolerating the state until the attentional system settles onto something available, is precisely the discomfort whose threshold is falling.

6.5 Effort is costly and effort is valuable

Two literatures must be held together. Wouter Kool, Joseph McGuire and Matthew Botvinick, then at Princeton, established in “Decision Making and the Avoidance of Cognitive Demand” (*JEP: General*, 2010) what they termed the **law of least mental effort**: given free choice between two tasks differing only in demand, people reliably choose the less demanding one, and will accept meaningful costs to do so. Botvinick’s wider legacy includes the conflict-monitoring theory of anterior cingulate function (with Braver, Barch, Carter and Cohen, *Psychological Review*, 2001) and, latterly, the meta-reinforcement-learning work at DeepMind.

Against this, Michael Inzlicht at the University of Toronto — one of the most effective internal critics of the ego-depletion literature, and a principal figure in the reform of self-control research — with Amitai Shenhav at Brown and Christopher Olivola at Carnegie Mellon, set out “The Effort Paradox: Effort Is Both Costly and Valued” (*Trends in Cognitive Sciences*, 2018). The evidence assembled there is that effort expended on an outcome *increases* the value assigned to it: the effect appears in humans and other animals, in contrafreeloading (animals with free food available will work for food), in the labour-and-love effect on self-assembled goods, and in the observation that many voluntarily chosen activities are selected for their difficulty rather than despite it.

The two findings are compatible and jointly damaging to the comfort-maximising strategy. Effort is aversive in prospect and value-conferring in retrospect. An organism optimising over prospective aversiveness will therefore systematically construct a life containing nothing it values, and will do so by a sequence of individually correct-feeling decisions.

6.6 What contracts is the world, not only the person

There is a second cost, less often noticed. James J. Gibson’s ecological psychology (below, §11.1) treats the environment as furnished with **affordances**: possibilities for action specified relative to the perceiver’s capacities. Eleanor Gibson’s complementary work on perceptual learning (*Principles of Perceptual Learning and Development*, 1969) established that perceptual development is a process of **differentiation** — learning to detect distinctions and invariants that were always available in the stimulation but were not previously picked up.

Both halves have a consequence for avoidance. Affordances are relative to what the organism can do; as the repertoire contracts, the set of perceived affordances contracts with it, and it does so *perceptually*, not merely propositionally. The world genuinely comes to contain less. And since differentiation requires exposure, the person who avoids a class of situations does not merely fail to acquire skill in it; they fail to develop the perceptual capacity to detect what those situations offer. Avoidance is thus not a neutral withdrawal that leaves the option open. It degrades the perceptual basis on which the option would have to be recognised.

7 Comfort with nobody in it

The preceding section treats discomfort tolerance as declining. This section treats the substitute. The claim is specific: **self-administered comfort delivers the sensory signature of relief while omitting the variable that made relief adaptive, and therefore does not terminate the state it appears to address.**

7.1 Social regulation is the default; self-regulation is the expensive exception

James Coan, at the University of Virginia, working with Hillary Schaefer and Richard Davidson at the University of Wisconsin–Madison, ran the experiment that anchors this claim. “Lending a Hand: Social Regulation of the Neural Response to Threat” (*Psychological Science*, 2006) placed married women in a scanner under threat of electric shock in three conditions: holding their husband’s hand, holding an anonymous male experimenter’s hand, and alone. Threat-related activation in a network including hypothalamus, ventral anterior cingulate, right dorsolateral prefrontal cortex and the caudate was **attenuated by hand-holding, most by the spouse’s hand, and the magnitude of attenuation in several regions tracked self-reported marital quality.** Subjective unpleasantness followed the same ordering.

Coan and Lane Beckes generalised this into **social baseline theory** (“Social Baseline Theory: The Role of Social Proximity in Emotion and Economy of Action”, *Social and Personality Psychology Compass*, 2011). The theory’s central move is to treat proximity to reliable others as the human brain’s **assumed baseline condition** rather than as a bonus. Because the brain economises on action, and because risk and effort can be distributed across a group, the presence of trusted conspecifics lowers the perceived cost of demands in the environment. Self-regulation — managing threat alone — is on this account not the normal mode plus support, but the abnormal mode, entered when the baseline assumption fails, and metabolically expensive because it is.

If this is right, then a comfort that arrives without another person does not restore the baseline; it manages a state produced by the baseline’s absence. The soothing is symptomatic.

An adjacent and much-publicised finding should be handled with care. Simone Schnall, Kent Harber, Jeanine Stefanucci and Dennis Proffitt, at the University of Virginia, reported in “Social Support and the Perception of Geographical Slant” (*Journal of Experimental Social Psychology*, 2008) that hills were judged less steep when participants were accompanied by a friend, with the effect graded by relationship quality and duration. Proffitt’s embodied-perception programme, of which this is a part, has been the subject of a sustained methodological critique by Chaz Firestone at Johns Hopkins and Brian Scholl at Yale (“Cognition Does Not Affect Perception”, *Behavioral and Brain Sciences*, 2016), who argue that such results reflect demand characteristics, judgement rather than perception, and the confusion of attention with perception. The critique has considerable force. The Coan result, which measures neural and self-report responses rather than psychophysical judgements, does not depend on it.

7.2 The loneliness spiral is a self-image mechanism

John Cacioppo (1951–2018) at the University of Chicago was, with Gary Berntson, the founder of social neuroscience as a named discipline; his wider contributions include the Elaboration Likelihood Model of persuasion with Richard Petty, the development of psychophysiological methods for social psychology, and the multi-level analysis programme that made social variables tractable to neuroscience. His last major programme was loneliness.

The essential findings, summarised in Cacioppo and Louise Hawkley’s “Perceived Social Isolation and Cognition” (*Trends in Cognitive Sciences*, 2009), are these. Loneliness is **perceived** isolation, only loosely coupled to objective social network size. It induces implicit hypervigilance

for social threat: lonely individuals detect negative social information faster, remember more of it, and evaluate ambiguous social behaviour more negatively, without awareness of doing so. This produces behaviour — guardedness, reduced self-disclosure, pre-emptive withdrawal — that elicits the rejection it anticipates, and confirms the model.

The causal direction was established longitudinally. Cacioppo, Hawkley and Ronald Thisted, “Perceived Social Isolation Makes Me Sad” (*Psychology and Aging*, 2010), used five annual waves of the Chicago Health, Aging and Social Relations Study and cross-lagged panel analysis to show that loneliness predicts subsequent depressive symptomatology while depressive symptomatology does not predict subsequent loneliness, with the relation surviving controls for demographics, objective network characteristics, stress and social support.

The hard outcomes are established by Julianne Holt-Lunstad at Brigham Young University. Her meta-analysis with Timothy Smith and J. Bradley Layton, “Social Relationships and Mortality Risk” (*PLoS Medicine*, 2010), covering 148 studies and 308,849 individuals, found a 50% increased likelihood of survival for participants with stronger social relationships, an effect comparable in magnitude to established risk factors such as smoking and exceeding obesity and physical inactivity. Her 2015 meta-analysis in *Perspectives on Psychological Science* separated loneliness, objective isolation and living alone, and found all three to be significant predictors of mortality with comparable effect sizes. Sheldon Cohen at Carnegie Mellon had already shown the mechanism operating on a measurable biological endpoint: in “Social Ties and Susceptibility to the Common Cold” (*JAMA*, 1997), participants inoculated with rhinovirus under quarantine developed clinical colds at rates inversely related to the diversity of their social ties, in a dose-response relation, controlling for immune status and health behaviours.

7.3 Why self-supplied comfort does not close the loop

Three arguments, in increasing order of specificity.

First, the **set-point argument**. If Coan and Beckes are right that the regulatory system’s reference condition is conspecific proximity, then comfort that alters affect without altering proximity leaves the reference error in place. The subjective relief is real and the underlying discrepancy is untouched, which is why the comfort must be repeated on a schedule.

Second, the **information argument**. What is regulating in the hand-holding study is not tactile pressure; a stranger’s hand produced significantly less attenuation than a spouse’s, and marital quality moderated it. The active variable is not sensation but what the sensation licenses the perceiver to infer about their standing. Harry Reis at the University of Rochester, with Margaret Clark and John Holmes, made this the organising construct of the intimacy literature under the name **perceived partner responsiveness**: the benefit of support depends on the recipient’s inference that the other party *understood, valued and cared for* the specific self that was disclosed. A self-administered comfort makes no such inference available, because there is no other party whose behaviour could carry the information.

Third, the **hypervigilance argument**. Cacioppo’s model implies that self-supplied comfort during loneliness is worse than neutral: it reduces the aversive signal that would otherwise motivate social approach, while leaving the implicit threat-detection bias in place and unexercised. The signal is silenced; the condition it signalled continues; and the corrective behaviour is not emitted.

7.4 The necessary distinction: solitude is not isolation

The argument above is not an argument against being alone, and the distinction is empirically supported rather than merely diplomatic. Thuy-vy Nguyen, then at the University of Rochester and now at Durham University, with Richard Ryan and Edward Deci, reported in “Solitude as an Approach to Affective Self-Regulation” (*Personality and Social Psychology Bulletin*, 2018)

that time spent alone reliably **deactivates** affect — reducing both high-arousal positive and high-arousal negative states — and that this deactivation effect is not itself unpleasant. Whether solitude is restorative or corrosive depended on autonomy: solitude chosen for its own sake predicted well-being, whereas solitude entered because social contact was unavailable or aversive did not.

The operative variable, then, is not the presence or absence of others but whether the state is entered as an approach or as an avoidance. That is exactly the distinction the practical section below is built to detect, and it cannot be made from the topography of the behaviour. Two people doing an identical thing alone on an identical evening may be doing opposite things.

8 Possession, latency, and a civilisation of abundance

A further mechanism deserves separate treatment, because it explains why the self-image in conditions of material surplus takes an acquisitive form.

8.1 The drive is mammalian and it is old

The **endowment effect** — the increase in an object’s valuation that follows from possessing it — was named by Richard Thaler (b. 1945), at the University of Chicago Booth School of Business and the 2017 Nobel laureate in economics, whose wider legacy includes mental accounting, the behavioural analysis of savings and pension defaults, and the choice-architecture programme. The definitive demonstration is Kahneman, Knetsch and Thaler, “Experimental Tests of the Endowment Effect and the Coase Theorem” (*Journal of Political Economy*, 1990): participants randomly given a university mug demanded roughly twice as much to part with it as unendowed participants would pay to acquire it, eliminating almost all predicted trade. Daniel Kahneman (1934–2024), at Hebrew University, British Columbia, Berkeley and Princeton, and the 2002 Nobel laureate, had supplied the theoretical basis with Amos Tversky (1937–1996) in prospect theory: value is defined over changes from a reference point, and losses loom larger than equivalent gains.

The behaviour is not a cultural artefact of market societies. Keith Chen, Venkat Lakshminarayanan and Laurie Santos, then at Yale, established a token economy with capuchin monkeys and reported in “How Basic Are Behavioral Biases? Evidence from Capuchin Monkey Trading Behavior” (*Journal of Political Economy*, 2006) that the monkeys’ choices displayed both reference dependence and loss aversion of roughly the human magnitude. Santos’s Comparative Cognition Laboratory has extended this line across several species. Sarah Brosnan, at Georgia State University, with Frans de Waal (1948–2024) of Emory and the Yerkes National Primate Research Center — whose wider legacy is the demonstration of reconciliation, empathy, fairness sensitivity and political behaviour in non-human primates — reported endowment effects in chimpanzees (*Current Biology*, 2007), with the interesting refinement that the effect was strongest for food items, the class of goods for which the underlying evolutionary logic is sharpest.

8.2 Latency, and the missing stop signal

Randolph Nesse, at the University of Michigan and subsequently founding director of the Center for Evolution and Medicine at Arizona State, established evolutionary medicine as a field with George C. Williams in *Why We Get Sick* (1994). Its central analytic move is the **mismatch** explanation: a trait that was adaptive in the environment where it evolved becomes pathogenic when the environment changes faster than the trait.

Possession behaviour has the classic mismatch structure, with one refinement that matters here. In scarcity, acquisition is terminated by external constraints — perishability, portability, the physical limits of what can be cached and defended. The drive requires no internal satiation

signal because the environment supplies one. Remove the constraint, and a drive that never needed a stop signal is revealed to lack one. This is the sense in which the possession drive is **latent**: it is not that it was absent and has appeared, but that it was always uncapped and this was invisible while the world did the capping.

Randy Frost at Smith College and Gail Steketee at Boston University established the pathological extreme and, with it, the diagnostic category: their cognitive-behavioural model of compulsive hoarding (*Behaviour Research and Therapy*, 1996) and three decades of subsequent work led to hoarding disorder’s recognition as an independent diagnosis in DSM-5 in 2013, separate from obsessive-compulsive disorder. Their model identifies information-processing deficits, beliefs about the nature of possessions, and emotional attachment to objects — including attachment to objects as parts of the self, and to objects as insurance against a future in which they might be needed.

8.3 The self-image as the possession that cannot be satiated

Russell Belk, at York University’s Schulich School of Business in Toronto and previously at Utah, made the connection explicit in “Possessions and the Extended Self” (*Journal of Consumer Research*, 1988), one of the most-cited papers in consumer research. Belk’s thesis is that people regard possessions as parts of themselves, that loss of possessions is experienced as a diminution of self, and that acquisition is therefore in part self-construction. His update, “Extended Self in a Digital World” (*JCR*, 2013), extends the analysis to profiles, avatars and accumulated digital traces, adding dematerialisation, co-construction of the self by others’ contributions, and the reconstruction of memory through curated archives.

Combine Belk with the mismatch argument and the result is the mechanism this section is for. When possession migrates from material objects to identity-constitutive abstractions — status, reputation, followers, credentials, the record of one’s own distinctiveness — every natural terminator is removed at once. Such goods do not perish, occupy no space, and cannot be exhausted. Worse, they are **positional**: their value is defined by relative standing, so that universal acquisition produces no aggregate gain, a structure analysed by Fred Hirsch in *Social Limits to Growth* (1976) and developed by the economist Robert Frank at Cornell in *Choosing the Right Pond* (1985) and *Luxury Fever* (1999).

Tim Kasser, at Knox College, supplied the psychological outcome data. With Richard Ryan he published “A Dark Side of the American Dream: Correlates of Financial Success as a Central Life Aspiration” (*JPSP*, 1993), showing that the relative centrality of extrinsic aspirations — wealth, fame, image — predicts lower well-being and higher distress, and that what matters is their weight relative to intrinsic aspirations (growth, relationships, community) rather than their absolute level. *The High Price of Materialism* (2002) summarises the subsequent programme, which includes longitudinal and experimental evidence and the finding that materialistic orientation is elevated by insecurity, including experimentally induced mortality salience and economic threat.

The synthesis: an uncapped mammalian acquisition system, released by abundance from its environmental terminators, finds in the self-image an object that is infinitely accumulable, socially graded, and never sufficient. The drive is not new. The object is.

9 The marketplace of emotional dependences

The self-image is self-oriented by construction, and a self-oriented agent transacting in comfort produces a market. This section argues that the market form destroys the good it trades, and that the destruction is mechanical rather than moral.

9.1 Two relational grammars, and what happens when they mix

Margaret Clark, at Carnegie Mellon and subsequently Yale, and Judson Mills, at the University of Maryland, drew the distinction that organises the modern literature. In **exchange relationships**, benefits are given with the expectation of comparable return, and a ledger is kept. In **communal relationships**, benefits are given in response to the other's needs, without obligation of specific repayment, and the keeping of a ledger is itself a violation.

The experimental result is counterintuitive and decisive. In Clark and Mills, "Interpersonal Attraction in Exchange and Communal Relationships" (*JPSP*, 1979), and in the programme of studies that followed, **repayment for a benefit decreased attraction when a communal relationship was desired and increased it when an exchange relationship was desired**. Being paid back is an insult in one grammar and a courtesy in the other. Their 1993 clarification ("The Difference Between Communal and Exchange Relationships: What It Is and Is Not") emphasises that the distinction is not about equality of outcomes or about the strength of the relationship but about the **rule governing the giving**.

Alan Page Fiske, at the University of California, Los Angeles, generalised the taxonomy in "The Four Elementary Forms of Sociality" (*Psychological Review*, 1992), an anthropologically grounded framework proposing that all social relations are constructed from four models: **communal sharing, authority ranking, equality matching** and **market pricing**. Fiske's central empirical claim is that people are highly sensitive to which model is in force, that errors of model are experienced as serious transgressions, and that the models are not interchangeable descriptions of the same underlying exchange.

Philip Tetlock, at Berkeley and later the Wharton School — whose wider legacy is the twenty-year forecasting-accuracy study published as *Expert Political Judgment* (2005) and the Good Judgment Project — demonstrated the affective force of model violation in "The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals" (*JPSP*, 2000). Merely contemplating an exchange that crosses from a communal to a market-pricing frame — pricing a body part, a child, a friendship — produced moral outrage, and, notably, **moral cleansing**: participants subsequently volunteered for unrelated altruistic acts, as if to expunge the contamination of having considered the trade. The mere entertainment of the market frame was experienced as a defilement requiring repair.

9.2 The price mechanism destroys what it prices

Uri Gneezy, at the University of California, San Diego's Rady School, and Aldo Rustichini at Minnesota, provided the field demonstration in "A Fine Is a Price" (*Journal of Legal Studies*, 2000). Israeli day-care centres introduced a monetary fine for parents collecting children late. Late collections **increased**, substantially, and did not return to baseline when the fine was withdrawn. The fine had reclassified the situation: what had been a communal obligation, discharged by consideration for the staff, became a priced service, purchasable. The reclassification was durable; removing the price did not restore the obligation.

The intra-psyche analogue is the overjustification effect. Edward Deci (b. 1942) and Richard Ryan (b. 1953), at the University of Rochester, are the authors of **self-determination theory**, the most extensively validated contemporary theory of motivation, organised around the basic needs for autonomy, competence and relatedness. Deci's original experiments (*JPSP*, 1971) showed that paying people to perform an intrinsically interesting activity reduced their subsequent voluntary engagement with it, and the meta-analysis by Deci, Koestner and Ryan (*Psychological Bulletin*, 1999), covering 128 experiments, established that tangible, expected, performance-contingent rewards reliably undermine intrinsic motivation. Self-determination theory's account of self-esteem is directly relevant: **contingent self-esteem** is classified as an *introjected* regulation — an internalised but unassimilated demand, experienced as pressure from within — and is associated

with poorer persistence, greater anxiety and worse well-being than autonomous regulation, even when it produces comparable short-term output.

An adjacent finding should be flagged as contested rather than relied on. Kathleen Vohs, at the University of Minnesota’s Carlson School, with Nicole Mead and Miranda Goode, reported in “The Psychological Consequences of Money” (*Science*, 2006) that subtle reminders of money made participants more self-sufficient — less likely to ask for help, less likely to give it, and inclined to place more physical distance between themselves and a stranger. The result is thematically ideal for this section and its replication record is poor: Rohrer, Pashler and Harris (*JEP: General*, 2015) failed to replicate related money-priming effects, subsequent meta-analytic work has found small effects with clear evidence of publication bias, and money priming is now a standard example in discussions of the replication crisis. It is cited here as a hypothesis with an attractive shape and inadequate support, which is how it should be treated.

9.3 The mechanism that spoils the soothing: attributional discounting

The precise reason a market in emotional support degrades its product was worked out before the market existed, by Harold Kelley (1921–2003) at the University of California, Los Angeles. Kelley’s wider legacy is interdependence theory, developed with John Thibaut, which analysed relationships in terms of outcome matrices and transformations of given situations, and the covariation model of causal attribution. The relevant principle is stated in “The Processes of Causal Attribution” (*American Psychologist*, 1973): the **discounting principle** holds that the role of a given cause in producing an effect is discounted if other plausible causes are also present.

Apply this to being comforted. The regulatory value of another person’s care, as §7.3 argued, lies in the inference it licenses: *this person attended to my state because of something about me and my standing with them*. That inference requires that no sufficient alternative cause be available. Introduce payment, obligation, reciprocal debt, professional role, algorithmic selection, or a reputational return for the giver, and a sufficient alternative cause is present. The discounting principle then operates automatically and without deliberation: the care is attributed to the incentive, the inference about standing is unavailable, and **the comfort is received as sensation without information**.

This is not a claim that paid or professional care is worthless; it is a claim about which component is doing which work, and it explains an otherwise puzzling regularity — that support can be abundant, competent, well-intentioned and unsatisfying, and that the recipient cannot say what is missing because what is missing is not in the support but in what the support permits them to conclude.

The self-image is what generates the market. An agent whose primary object is the maintenance of a favourable self-model necessarily approaches others as inputs to that model: sources of validation, audiences, comparison classes, and, at best, exchange partners in a mutual validation arrangement. Such an agent cannot generate the communal grammar, because the communal grammar requires attending to the other’s state as the terminal object rather than as an instrument. And the market it constructs is one in which every participant knows there is a sufficient alternative cause for every kindness received.

10 Affirmation hunger and the tacit contract of the profile

10.1 Why the craving is structural, not weak-willed

Assemble the pieces already laid out. Self-esteem is a gauge of relational value (Leary). It is contingent on performance in staked domains (Crocker). When it is high and fragile it must be actively defended (Kernis). It is transparent to its owner, who experiences its readings as facts

about the world rather than as outputs of a monitoring system (Metzinger). It is verified rather than optimised, so that even unflattering content is protected (Swann). And its maintenance is now conducted in an environment where the relevant evidence arrives as a discrete, countable, socially aggregated stream.

The result is a system that requires continuous external input to maintain a state it treats as intrinsic. The word for that in the operant literature is not weakness. It is a **schedule**. Charles Ferster and B. F. Skinner (1904–1990), at Harvard, established in *Schedules of Reinforcement* (1957) that variable-ratio schedules — reinforcement delivered after an unpredictable number of responses — generate the highest and most persistent response rates and the greatest resistance to extinction of any simple schedule. Social feedback on a distributed platform is a variable-ratio schedule with a short response cost, and the behaviour it maintains has the properties the schedule predicts: high rate, insensitivity to satiation, resistance to extinction, and persistence long after the reinforcer has ceased to be enjoyed. The person checking is not failing to notice that it is unrewarding. The schedule does not require that they enjoy it.

10.2 The tacit contract: how profiles are made real

The mechanism by which a manifestly constructed self-presentation acquires the status of a fact was described in full before any of the relevant technology existed.

Erving Goffman (1922–1982) — Canadian-born, trained at Chicago, at Berkeley and then the Benjamin Franklin Professor of Anthropology and Sociology at the University of Pennsylvania, president-elect of the American Sociological Association at his death — is the analyst of interaction as an order in its own right. His wider legacy includes *Asylums* (1961) on total institutions, *Stigma* (1963) on spoiled identity, *Frame Analysis* (1974) on the organisation of experience, and the essay “On Face-Work” (*Psychiatry*, 1955).

The Presentation of Self in Everyday Life (1959) supplies the contract. Goffman’s argument is that in any encounter each participant projects a definition of the situation and a claim about the kind of person they are, and that the others do not, in the ordinary case, test that claim. They **accede** to it. The result is a *working consensus*: an agreement not about what is true but about **whose claims will go unchallenged**. Crucially, Goffman identifies the reason for the forbearance. It is not politeness or credulity. Each participant’s own presentation is equally vulnerable, and a challenge to one licenses challenges to all. The contract is mutual non-aggression among performers, maintained because everyone present has an exposure. Breaches — Goffman’s *incidents* — are met with rapid collaborative repair by all parties, including those the breach did not concern, which is the clearest evidence that the interest being protected is the order itself rather than any individual’s face.

To this Goffman adds **audience segregation**: performances are kept coherent by ensuring that the audiences for incompatible presentations do not meet. That safeguard is exactly what networked profiles remove. Alice Marwick, now at the University of North Carolina at Chapel Hill, and danah boyd, founder of the Data & Society Research Institute, named the consequence **context collapse** in “I Tweet Honestly, I Tweet Passionately” (*New Media & Society*, 2011): a single utterance is addressed to an *imagined audience* while being available to every audience simultaneously, so the performer must construct a lowest-common-denominator self that no context would have called for. The profile is thereby not merely a performance but a performance optimised against an audience that does not exist.

The second half of the contract is supplied by **pluralistic ignorance**, formalised in the modern literature by Deborah Prentice, then at Princeton and subsequently Vice-Chancellor of the University of Cambridge, and Dale Miller at Stanford. In “Pluralistic Ignorance and Alcohol Use on Campus” (*JPSP*, 1993) they showed that students systematically overestimated their peers’ comfort with campus drinking practices, because each inferred others’ private attitudes from others’ public compliance, while knowing their own compliance to be insincere. Every

individual privately dissented; each believed themselves alone in dissenting; and the collective behaviour was maintained by the aggregation of people acting against their own judgement.

Together these give the full account of how a profile becomes real. Each participant knows their own profile to be curated. Each observes that others' profiles are treated as informative. Each infers that others do not share their scepticism. Each therefore treats others' profiles as informative, both to avoid the social cost of scepticism and because to attack the genre would expose their own entry in it. No one is deceived, no one believes, and everyone behaves as though everyone believes. Goffman supplies the motive for the forbearance; Prentice and Miller supply the inferential error that makes it stable; Ferster and Skinner supply the schedule that keeps people supplying material to it.

10.3 Calibration: what the evidence on digital media does and does not show

An essay of this kind is under obligation not to accept an agreeable conclusion on weak data. The evidence on digital media and well-being is weaker than the surrounding discourse assumes, and the relevant literature says something more precise than “screens are bad”.

Amy Orben, now at the Medical Research Council Cognition and Brain Sciences Unit in Cambridge, and Andrew Przybylski at the Oxford Internet Institute, applied **specification curve analysis** — computing the effect across every defensible analytic specification rather than reporting one — to three large representative datasets in “The Association Between Adolescent Well-Being and Digital Technology Use” (*Nature Human Behaviour*, 2019). Across 20,000-plus specifications the median association was negative but tiny, accounting for well under half a percent of variance in well-being, and comparable in magnitude to associations with eating potatoes or wearing spectacles. Their wider contribution is methodological: they demonstrated that the field's headline findings were artefacts of specification choice, and that a determined analyst could produce effects of either sign from the same data.

That is the correct baseline. Against it, two more specific findings survive.

Ethan Kross, at the University of Michigan, where he directs the Emotion and Self-Control Laboratory, used experience sampling rather than aggregate self-report in Kross et al., “Facebook Use Predicts Declines in Subjective Well-Being in Young Adults” (*PLoS ONE*, 2013): text-messaged five times daily over two weeks, participants' Facebook use at one prompt predicted worse affect at the next and declines in life satisfaction over the fortnight, whereas direct social interaction predicted improvement. The **within-person, time-lagged** design is what makes this informative where cross-sectional correlations are not. Philippe Verduyn, Kross and colleagues then isolated the moderator (*JEP: General*, 2015): **passive** use — consuming others' curated material without interacting — undermined affective well-being, mediated by envy, while **active** use did not.

The pattern that emerges is consistent with this essay's thesis and inconsistent with the simpler moral panic. The medium is not the variable. The variable is whether the activity places the user in the position of an evaluator comparing a self-image against a stream of others' self-images, with no action available. That is a self-referential processing task with no motor output, which is to say it is rumination with an external feed.

Kross's wider work on inner speech, summarised in *Chatter* (2021), supplies the tools used in §12 below.

11 The active mind: readiness, absorption, and why release is hard

The preceding sections were negative. This one specifies the positive condition the self-image displaces, and then explains why displacing it back is difficult.

11.1 Perception is for action, and it runs at the world's tempo

James J. Gibson (1904–1979) spent his career at Smith College and then Cornell, and produced across three books — *The Perception of the Visual World* (1950), *The Senses Considered as Perceptual Systems* (1966) and *The Ecological Approach to Visual Perception* (1979) — the most sustained alternative to the inferential account of perception. His wartime work on pilot selection and landing had convinced him that the laboratory's static, monocular, impoverished stimulus was not the stimulus organisms actually encounter. His mature position holds that the optic array available to a moving observer contains **invariants** that specify the layout and the observer's own motion through it directly, without inferential reconstruction; that the appropriate unit of analysis is the **animal-environment system**, not the animal; and that what perception delivers is **affordances** — the possibilities for action that a surface, object or arrangement offers *this* organism with *these* capacities. A surface affords walking-on for a creature of the right size and posture and does not for another. Affordances are neither objective properties nor subjective impositions; they are relational facts, and they are perceived.

Eleanor J. Gibson (1910–2002), also at Cornell — whose visual cliff experiments with Richard Walk are among the most reproduced findings in developmental psychology, and who received the National Medal of Science in 1992 — supplied the learning theory. In *Principles of Perceptual Learning and Development* (1969) she argued that perceptual learning is not the association of meanings to sensations but **differentiation**: the progressive detection of information that was present all along. Expertise is a change in what one can pick up.

Two consequences for this essay. First, if perception is *for* the guidance of action and is specified continuously by a moving observer's own movement, then perception has a tempo, and it is the tempo of the environment. It cannot be time-shared with a slow verbal process without loss (§2.3). Second, if affordances are relative to capacities and are detected through differentiation acquired by exposure, then what one is able to perceive is a function of what one has done. The self-image, which specifies what one does, is thereby also a constraint on what one can see.

11.2 Enaction: the perceiver is not a spectator

Francisco Varela (1946–2001), the Chilean biologist who with Humberto Maturana developed the theory of **autopoiesis** in the early 1970s, and who spent his later career at the CNRS and the LENA laboratory at the Hôpital de la Salpêtrière in Paris, is the principal architect of the enactive approach. His wider legacy includes the founding of the Mind and Life Institute and the programme of **neuropsychophenomenology** (“Neuropsychophenomenology: A Methodological Remedy for the Hard Problem”, *Journal of Consciousness Studies*, 1996), which proposed disciplined first-person reports, gathered from trained observers, as constraints on neural models rather than as data to be explained away.

The Embodied Mind (1991), with Evan Thompson — now at the University of British Columbia and author of *Mind in Life* (2007) — and Eleanor Rosch at Berkeley, whose prototype theory of categorisation had already dismantled the classical definitional account of concepts, sets out the position: cognition is **enacted**, brought forth through the sensorimotor coupling of an autonomous system with its surroundings, rather than computed over representations of a pre-given world.

The relevance is the account of absorption. On an enactive reading, being fully present is not

a special state added to ordinary perception; it is ordinary perception operating without the interposition of a represented self. The represented self is the thing that converts participation into spectatorship.

Hubert Dreyfus (1929–2017), at the University of California, Berkeley, made the same point in the phenomenological register and made a career of it. His *What Computers Can't Do* (1972) was the earliest and, in retrospect, most accurate critique of symbolic artificial intelligence; the five-stage skill model developed with his brother Stuart in *Mind Over Machine* (1986) — novice, advanced beginner, competent, proficient, expert — described the progressive disappearance of rule-following in favour of situational discrimination. His term for the expert's mode is **absorbed coping**: the skilled agent responds to the solicitations of the situation without representing goals, alternatives or themselves. His long exchange with John McDowell (*Inquiry*, 2007) over whether conceptual rationality is present throughout skilled action, or whether it must be suspended for absorption to occur, remains unresolved and is a genuine philosophical dispute rather than an empirical one; but the descriptive claim that expert performance is disrupted by the reintroduction of explicit self-monitoring is the same claim Beilock demonstrated in the laboratory (§4.1).

11.3 Why the action mind competes with perception

Three convergent literatures explain why the intruding process is so hard to shut down: it uses the same machinery as the thing it interrupts.

Motor imagery. Marc Jeannerod (1935–2011), at the Institut des Sciences Cognitives in Lyon, founded the field of motor cognition. His work established that imagining an action recruits substantially the same cortical and subcortical structures as executing it, that imagined movements obey the same temporal laws as real ones — including Fitts's speed-accuracy trade-off — and that imagined effort produces autonomic changes proportional to the imagined load (“The Representing Brain”, *Behavioral and Brain Sciences*, 1994; “Neural Simulation of Action”, *NeuroImage*, 2001). The simulation of action is action machinery run offline.

Common coding. Wolfgang Prinz, founding director of the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig and previously of the Max Planck Institute for Psychological Research in Munich, argued that perception and action share a representational medium coded in terms of distal events (*European Journal of Cognitive Psychology*, 1997), reviving the nineteenth-century ideomotor principle within experimental psychology. If planned actions and perceived events are coded in the same format, then imagining and perceiving compete for the same representational resources.

Visual imagery. The oldest demonstration is Cheves West Perky's (1910): participants asked to imagine an object while a faint image of that object was projected below threshold incorporated the percept into their imagery and denied seeing anything. The contemporary treatment is Nadine Dijkstra and Stephen Fleming at University College London, whose work on **perceptual reality monitoring** (see Dijkstra, Bosch and van Gerven, *Trends in Cognitive Sciences*, 2019; Dijkstra and Fleming, *Nature Communications*, 2023) shows substantial overlap between the neural substrates of imagery and perception and that the discrimination between them is made by a threshold on signal strength rather than by a categorical tag. Imagination is not a separate channel. It is the same channel, driven from inside.

Taken together: the self-image in its active form — the rehearsed conversation, the imagined observer, the anticipated evaluation — is not idle chatter alongside perception. It is the perceptual and motor system being driven by internal input, and every unit of capacity so occupied is unavailable to the world.

Perceptual decoupling is the measured consequence (Smallwood and Schooler, §4.3): when internal content takes over, the cortical response to sensory input is attenuated, and the subject does not notice. The functional-imaging counterpart is the anticorrelation between the default mode network and the dorsal attention network reported by Michael Fox, Maurizio Corbetta,

Marcus Raichle and colleagues (*PNAS*, 2005) — regions supporting self-referential and externally directed processing show opposed spontaneous fluctuations. The strength and even the existence of the anticorrelation depends on preprocessing choices, particularly global signal regression, which is a live methodological dispute; the behavioural evidence for the trade-off does not depend on resolving it.

11.4 The load principle, and the design implication

Nilli Lavie, at University College London’s Institute of Cognitive Neuroscience, established the variable that turns all of this into practical advice. Her **load theory** (“Perceptual Load as a Necessary Condition for Selective Attention”, *JEP: Human Perception and Performance*, 1995; Lavie, Hirst, de Fockert and Viding, *JEP: General*, 2004) resolves a decades-old dispute between early- and late-selection accounts of attention by making selection contingent on load. Under **high perceptual load**, capacity is exhausted by the relevant stimuli and irrelevant material is not processed. Under **low perceptual load**, spare capacity spills over and irrelevant material is processed whether or not the subject wishes it. High **cognitive** load, by contrast — occupied working memory — *increases* distractor interference, because the control mechanism that maintains task priority is itself loaded.

The implication is direct and unusually actionable. Self-referential content is a distractor of the second kind: it loads working memory rather than perception. It is therefore not suppressed by trying harder, which loads control further, but by raising **perceptual** load — by placing oneself in an environment whose sensory demands are high enough to exhaust capacity, and whose tempo is set externally.

This is why the recommendations in the next section are environmental rather than doxastic. One does not release the self-image by deciding to. One occupies the machinery it needs.

11.5 A remark on presence

Anil Seth, at the University of Sussex, where he co-directs the Sackler Centre for Consciousness Science, has developed the predictive-processing account of selfhood furthest in the interoceptive direction. His proposal, in Seth, Suzuki and Critchley’s “An Interoceptive Predictive Coding Model of Conscious Presence” (*Frontiers in Psychology*, 2012) and in *Being You* (2021), is that the sense of **presence** — of the world and oneself being really here — is not a perception of anything but the signature of successful predictive modelling of interoceptive and exteroceptive signals together; and that the experienced self is a controlled hallucination whose function is physiological regulation, not accurate self-portrayal.

If that is even approximately right, then the demand that the self-image be *accurate* is a category error twice over: it is not a description, and it was never built to be one. What can sensibly be asked of it is that it not be granted precision it has not earned, and that it not be permitted to run when something better is available to run instead.

12 Practical section: detection and audit

What follows is a method, not an exhortation. It has two parts: recognising self-image content in real time, and auditing choices after the fact. Both are designed to be usable without special conditions and without believing anything in particular.

12.1 How to detect a self-image thought

Self-image content has reliable formal markers, and the markers are easier to detect than the content is, because they do not require agreeing or disagreeing with what the thought says. Run the checks on form.

1. The copula test. Self-image thoughts are predominantly of the form *I am X, I am not X, I am the kind of person who*. Task thoughts are of the form *this needs, that is loose, next*. A thought whose main verb is *to be*, with oneself as subject, is almost always self-image. A thought whose main verb is transitive and has an object in the environment is almost always not.

2. The tense test. Perception is indexed to now. Self-image content is retrospective, prospective, or tenseless: *should have, always, never, by now*. A thought without a temporal index to the present situation is not about the present situation.

3. The comparative test. Look for the comparative and superlative — *better, worse, still not, enough, behind*. Comparison requires a standard, a standard requires a reference class, and a reference class is an audience. If a comparison is present, an audience is present.

4. The observer-perspective test. This is the sharpest single check, and it comes from clinical work on social anxiety (Hackmann, Clark and McManus, §4.5) and from the field-versus-observer distinction in autobiographical memory established by Georgia Nigro and Ulric Neisser (“Point of View in Personal Memories”, *Cognitive Psychology*, 1983). Ask: **in the mental image accompanying this thought, where is the viewpoint?** If you are seeing the scene from behind your own eyes, the content is situational. If you are seeing *yourself*, from outside, as a camera or another person would, the content is self-image. The check takes under a second and its answer is not negotiable by argument.

5. The terminal-state test. Ask what would count as this thought finishing. A task thought terminates in an action or a piece of information. A self-image thought terminates in a **verdict** — a rating, a settled judgement about what kind of person one is. If the thought’s completion condition is a rating, it is self-image, however much it presents itself as analysis.

6. The falsifiability test. Ask what observation would settle it. Self-image propositions are constructed so that no available observation settles them, which is why they can run indefinitely without progress. Genuine problems consume evidence and change shape. Self-image content is evidentially inert.

7. The load test. Increase the demands of what you are actually doing — take a harder line, a faster tempo, a heavier object. Task-relevant thought scales with the task. Self-image content is *displaced* by load rather than intensified by it (§11.4). If a train of thought survives unchanged through a change of task difficulty, it was not about the task.

8. The linguistic marker, with its caveat. James Pennebaker, at the University of Texas at Austin — originator of the expressive-writing paradigm (with Sandra Beall, 1986) and of the Linguistic Inquiry and Word Count text-analysis programme, which established that function words rather than content words carry most psychological signal — documented elevated first-person-singular pronoun use as a marker of self-focus. The careful version of the finding is Tackman, Sbarra, Mehl, Pennebaker and colleagues’ multi-laboratory synthesis (*JPSP*, 2019), which found that “I-talk” is more robustly associated with general **negative emotionality** than with depression specifically. Used as a within-person signal rather than a between-person diagnosis, it is serviceable: a marked rise in one’s own first-person density, in speech or writing, indicates that attention has moved onto the self, and it can be noticed by a listener before it is noticed by the speaker.

A note on what not to do with a detection. Do not argue with the content. Arguing with a self-image thought is a self-image thought, requires the same machinery, and confirms the premise that the question is important. The correct response to a detection is a change of task, not a rebuttal. This is the operational content of Hayes’s **defusion** and of Wells’s **detached mindfulness**: to alter one’s relation to the thought rather than its content.

12.2 How to audit a choice: eight questions

The distinction wanted here — self-isolating and self-comforting versus social, ecological and peer-comforting — cannot be made from the outward form of the activity, because the same act belongs to either class depending on its function (§7.4). Function is determined by what the act is maintained by and what it terminates in. The following questions determine that.

1. The counterparty question. *At the end of this, is another person present, and did their behaviour toward me depend on my particular state?* This is the operational form of social baseline theory (§7.1). Comfort with no counterparty does not reset the reference condition; it manages the error signal generated by its absence.

2. The discounting question. *If someone was warm to me here, is there a sufficient alternative explanation for their warmth?* Payment, professional role, obligation, algorithmic delivery, reciprocal debt, or an audience for their own performance all qualify. Where a sufficient alternative cause exists, Kelley’s discounting principle (§9.3) removes the inference that carries the benefit, and the episode will not satisfy however competently it is performed. This question predicts the aftertaste in advance.

3. The grammar question. *Am I keeping a ledger?* If you are tracking what you have given and what is owed, you are operating a market-pricing model (Fiske) in a relationship that requires communal sharing (Clark and Mills). The ledger is the diagnostic, not its balance.

4. The repertoire question. *Does this choice widen or narrow the set of situations I can enter tomorrow?* Avoidance is negatively reinforced and therefore feels correct at the moment of choice; the cost is deferred and invisible. Narrowing choices do not announce themselves as narrowing. Ask about the set, not the evening.

5. The decay question. *Would the discomfort I am removing have decayed on its own?* Most acute aversive states habituate within a period that is short relative to how long they feel. If the answer is yes, the act is not a solution but the acquisition of a dependency, and it lowers the next threshold (§6.2).

6. The tempo question. *Who sets the pace?* The self-image requires an unbounded clock; it cannot run in situations that make demands at a rate it cannot interrupt. Activities whose tempo is externally imposed — playing music with other people, contact sport, conversation with someone who will notice absence, physical work with a real deadline set by material rather than by policy — are structurally protective. Activities whose tempo is under your control (scrolling, most solitary consumption, most planning) are structurally hospitable to it.

7. The residue question. *When this is over, does anything exist that did not exist before?* An object, a repaired thing, a piece of work, a changed state in another person. This is not a productivity criterion. It is a test of whether the activity was coupled to anything outside the self-model at all.

8. The autonomy question. *Am I entering this state because I want it, or because the alternative is aversive?* Nguyen, Ryan and Deci (§7.4) found this to be the variable that determines whether solitude restores or depletes. It is the same act either way. The question is whether it is an approach or an escape, and the honest answer is usually available within a second of asking, before the justification arrives.

Instrumentation. These questions are more reliable if the data are collected at the time rather than reconstructed. The method is experience sampling, developed by Mihaly Csikszentmihalyi and Reed Larson at Chicago and validated in “Validity and Reliability of the Experience-Sampling Method” (*Journal of Nervous and Mental Disease*, 1987) — the technique that produced the flow literature and, later, Killingsworth and Gilbert’s mind-wandering result. Applied to oneself: several random prompts a day, each answered in under fifteen seconds with four items — *what am I doing; what is in front of me; who is present; is there an observer in the picture*. A fortnight of this produces a record that is not subject to the reconstructive distortions the self-image imposes on retrospective self-report, and it is normally the first time a person sees the actual distribution

of their days rather than the narrative summary of them.

12.3 Countermeasures ranked by evidential support

Strongest evidence: change the behaviour, leave the narrative alone. Behavioural activation (§3.3) is the best-supported intervention relevant to this essay’s thesis, and it works by scheduling contact with reinforcing activity and by functional analysis of avoidance patterns, without addressing self-beliefs. This is not a metaphor for the essay’s argument; it is the argument, tested against an active control and a drug.

Strong evidence: external focus of attention. Wulf’s manipulation (§4.2) is the cheapest intervention in this document. Reformulate any instruction you give yourself in terms of an effect in the environment rather than a state of your body or your person. The effect appears in accuracy, economy and retention, in novices and experts and patients.

Strong evidence: expectancy violation under tolerated discomfort. Michelle Craske, at the University of California, Los Angeles, where she directs the Anxiety and Depression Research Center — whose wider contributions include the development of standard treatment protocols for panic and generalised anxiety and work on the DSM anxiety criteria — reoriented exposure therapy in “Maximizing Exposure Therapy: An Inhibitory Learning Approach” (*Behaviour Research and Therapy*, 2014). The important finding for a lay reader is that the mechanism of benefit is **not** habituation, and within-session fear reduction does not predict outcome. What predicts outcome is **expectancy violation**: the discrepancy between what the person predicted would happen and what did. The practical corollary is that tolerating discomfort works when you have stated in advance, specifically, what you expect to go wrong, and then find out. Endurance without a prediction teaches much less.

Good evidence: self-distancing for unavoidable reflection. Some reflection is necessary. Ethan Kross and Ozlem Ayduk, at Berkeley, established that its form determines its effect. In Ayduk and Kross (“From a Distance”, *JPSP*, 2010) and Kross et al. (“Self-Talk as a Regulatory Mechanism: How You Do It Matters”, *JPSP*, 2014), analysing a distressing experience from a **self-distanced** perspective — using one’s own name or the second person, or adopting the position of an observer — produced construal rather than recounting, less emotional reactivity, lower cardiovascular reactivity, less rumination at follow-up, and better performance under social stress, relative to the ordinary first-person immersed mode. Note carefully that this is the *opposite* of the observer perspective in §12.1’s fourth test: the distinction is that here the object is a specific past or future *event* to be construed, whereas the pathological observer perspective takes the *self* as the object and produces no construal at all.

Moderate evidence: trained experiential focus. The Farb and Segal dissociation (§4.8), and the MBCT relapse-prevention trials on which Segal’s reputation rests, support attention training that separates experiential from narrative self-reference. The evidence for mindfulness-based interventions is strongest for depressive relapse prevention in patients with multiple prior episodes and considerably weaker elsewhere; the literature as a whole has well-documented problems with active control conditions and publication bias, and should not be cited as if it were uniform.

Moderate evidence, narrower base: metacognitive attention training. Wells’s Attention Training Technique and detached mindfulness (§4.5) target the process directly and have promising trial data from a small number of groups. Treat as promising rather than established.

Structural, no trial required: remove the conditions. From §11.4 and the tempo question: raise perceptual load, hand the clock to someone else, and reduce the number of contexts per day in which you are invited to summarise yourself. The last is the most neglected. Every request to describe oneself — a profile, a bio, an update, a rating, a status — is a rehearsal of the conceptual self, and rehearsal strengthens it. The frequency of these requests in contemporary life is historically unprecedented, and it is largely optional.

13 Limits, counter-evidence and what is not being claimed

An argument of this shape attracts overstatement, and several of the obvious overstatements are false.

The self-concept is not eliminable, and some of it is load-bearing. Hazel Rose Markus, at Michigan and then Stanford, where she co-founded the Center for Comparative Studies in Race and Ethnicity, established with Paula Nurius the construct of **possible selves** (*American Psychologist*, 1986): representations of what one might become, which function as incentives and as organising frameworks for behaviour, and whose absence in a domain predicts absence of effort in it. A self-representation oriented to the future and to action is not the object of this essay’s criticism; the object is the evaluative, comparative, audience-facing summary. Markus’s later work with Shinobu Kitayama (“Culture and the Self”, *Psychological Review*, 1991) adds the necessary cultural caveat: the independent self-construal that the foregoing analysis treats as the default is not universal, and the distribution and expression of the harms described here vary with construal. An essay written from within an interdependent construal would have had to locate the same pathologies in the relational rather than the individual self.

Broadening the self-concept genuinely buffers threat. Claude Steele’s self-affirmation theory, and the substantial intervention literature reviewed by Geoffrey Cohen at Stanford and David Sherman at Santa Barbara (*Annual Review of Psychology*, 2014), show that affirming values in one domain reduces defensiveness about threats in another, sometimes with effects that persist for years in field settings. This is a real counterweight — but note its mechanism, which is to *de-concentrate* the self-image so that no single domain is load-bearing. That is Crocker’s contingency analysis run in reverse, and it is consistent with the present argument rather than an exception to it: what is being reduced is stakes, not improved is content.

A different self-relation is available and better-evidenced than abolition. Kristin Neff, at the University of Texas at Austin, defined and operationalised **self-compassion** (*Self and Identity*, 2003) as self-kindness, common humanity and mindfulness, explicitly contrasting it with self-esteem on the grounds that it requires no evaluation, no comparison and no domain in which to succeed. The subsequent literature — with the usual caveats about self-report measurement and the ongoing dispute over the factor structure of her scale — associates it with lower anxiety and depression and greater resilience without the defensive correlates of high self-esteem. For a reader who finds the wholesale case against self-relation implausible, this is the well-supported middle position: keep the relation, remove the scoreboard.

Do not build on ego depletion. The intuitive model in which willpower is a depletable fuel is not adequately supported; the multi-laboratory registered replication (Hagger et al., 2016) and subsequent meta-analytic corrections for publication bias have left the effect in serious doubt, and Inzlicht among others has proposed motivational and attentional-shift accounts in its place. §6 was therefore constructed from opponent-process and allostatic mechanisms and from reinforcement contingencies, not from a resource model, and the reader should be suspicious of any version of this essay’s argument that runs on the tank metaphor.

Several cited findings are smaller than their reputations. Stated in the text where they occur, and repeated here: the field magnitude of stereotype threat is modest and heterogeneous; verbal overshadowing replicates at about a third of the original effect; money priming has largely failed to replicate; top-down effects on perception of the geographical-slant kind are subject to a serious methodological challenge; transient hypofrontality is not supported in its global form, only in the narrower selective form; the improvisation-imaging study had six participants; the free energy principle is a framework whose empirical commitments are contested.

The direction of causation is not always established. The loneliness literature has good longitudinal evidence for the direction argued (Cacioppo, Hawkley and Thisted, 2010) and the mind-wandering literature has good time-lagged evidence (Killingsworth and Gilbert, 2010). Others do not: the association between self-image preoccupation and a narrowed life is compatible

with the narrowing coming first.

What is not claimed. Not that a self does not exist; the ecological and interpersonal selves of Neisser's taxonomy are constitutive of perception and interaction. Not that reflection is worthless; §12.3 gives the conditions under which it helps. Not that solitude is harmful; §7.4 gives the moderator. Not that comfort is illegitimate; the argument is about which comforts terminate the state they address. Not that the self-image can be removed by deciding to remove it; the entire practical section is built on the assumption that it cannot, and that the workable interventions are environmental, attentional and behavioural.

The claim is narrower and, being narrower, harder to escape: **the narrative self is a high-latency, low-bandwidth, evidentially inert controller with a systematic bias toward verdicts, and nearly everything worth doing runs faster than it.**

14 Conclusion: the interruption you want and the interruption you get

Return to the two grammars.

The interruption in Reading A — action interrupting the self-image — has no mystery in it. It is what happens when a system with finite capacity is given a task whose demands exceed what is left over after the world is tracked. It has been produced in a scanner with difficulty-matched arithmetic, in a laboratory with a ski simulator and an instruction about where to look, in professional pianists on a plastic keyboard, and in every workshop, kitchen, rehearsal room and mountainside where the tempo belongs to something other than the person. It is not rare, it is not an achievement of character, and it is not the property of any tradition. It is a resource allocation.

The interruption in Reading B — the self-image interrupting action — is likewise unmysterious, and it is the ordinary condition. Its ingredients have been listed: a monitoring system mistaken for a control system (§3); a set of compiled skills degraded by the reinstatement of explicit supervision (§4); settled preferences that direct search away from what would revise them, while the search feels open (§5); a threshold for discomfort that falls with each avoidance because the avoidance is negatively reinforced and the set point drifts (§6); a supply of comfort that omits the only variable that made comfort adaptive, and therefore has to be repeated (§7); an uncapped acquisition system that has found in identity an object with no satiation point (§8); a market in emotional dependence in which every kindness has a sufficient alternative cause and is therefore discounted on arrival (§9); and a variable-ratio schedule of affirmation, sustained by a contract that nobody signed, nobody believes, and everybody honours because their own entry is at stake (§10).

None of these is a moral failing and none is corrected by insight. That is why §12 contains no advice to think differently about oneself. Thinking differently about oneself is more self-image, produced by the same machinery, on the same schedule, and it is the most reliable way to spend a decade making no change at all.

What can be done is structural, and it is short. The self-image requires four conditions: **an unbounded clock, an audience, low perceptual load, and a private ledger.** Each is removable, and removing any one degrades the process substantially.

Take the clock: enter activities whose tempo is set by material, by other people, or by physics, and which impose a cost for absence that arrives before you have noticed you left.

Take the audience: reduce the number of daily occasions on which you are asked to summarise yourself, and notice, using the observer-perspective test, when one has installed itself without invitation.

Take the load: put yourself in front of things that demand differentiation — a difficult text,

an instrument, an unfamiliar landscape, a person actually saying something — because capacity spent on perception is capacity the narrator cannot have.

Take the ledger: give without recording, receive without repaying on the spot, and be suspicious of every arrangement in which your comfort arrives with a sufficient alternative explanation attached.

What remains after that subtraction is not a better self-image. It is Neisser's ecological self, which was never in question, doing what it was built for: specifying where you are, what is in reach, and what the situation affords, continuously, at the speed of the situation, without an opinion about you.

References

- Alogna, V. K., et al. (2014). Registered Replication Report: Schooler & Engstler-Schooler (1990). *Perspectives on Psychological Science*, 9(5), 556–578.
- Ayduk, Ö., & Kross, E. (2010). From a distance: Implications of spontaneous self-distancing for adaptive self-reflection. *Journal of Personality and Social Psychology*, 98(5), 809–829.
- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, 46(3), 610–620.
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, 4(1), 1–44.
- Beckes, L., & Coan, J. A. (2011). Social baseline theory: The role of social proximity in emotion and economy of action. *Social and Personality Psychology Compass*, 5(12), 976–988.
- Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, 130(4), 701–725.
- Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive. *Journal of Experimental Psychology: Applied*, 8(1), 6–16.
- Belk, R. W. (1988). Possessions and the extended self. *Journal of Consumer Research*, 15(2), 139–168.
- Belk, R. W. (2013). Extended self in a digital world. *Journal of Consumer Research*, 40(3), 477–500.
- Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*, 6, 1–62.
- Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition*, 108(3), 652–661.
- Botvinick, M., & Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature*, 391, 756.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652.
- Brosnan, S. F., & de Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature*, 425, 297–299.
- Brosnan, S. F., et al. (2007). Endowment effects in chimpanzees. *Current Biology*, 17(19), 1704–1707.
- Cacioppo, J. T., & Hawkley, L. C. (2009). Perceived social isolation and cognition. *Trends in Cognitive Sciences*, 13(10), 447–454.
- Cacioppo, J. T., Hawkley, L. C., & Thisted, R. A. (2010). Perceived social isolation makes me sad: Five-year cross-lagged analyses. *Psychology and Aging*, 25(2), 453–463.
- Chen, M. K., Lakshminarayanan, V., & Santos, L. R. (2006). How basic are behavioral biases? Evidence from capuchin monkey trading behavior. *Journal of Political Economy*, 114(3), 517–537.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. Heimberg et al. (Eds.), *Social Phobia: Diagnosis, Assessment and Treatment*. Guilford.
- Clark, M. S., & Mills, J. (1979). Interpersonal attraction in exchange and communal relationships. *Journal of Personality and Social Psychology*, 37(1), 12–24.
- Clark, M. S., & Mills, J. (1993). The difference between communal and exchange relationships: What it is and is not. *Personality and Social Psychology Bulletin*, 19(6), 684–691.

- Coan, J. A., Schaefer, H. S., & Davidson, R. J. (2006). Lending a hand: Social regulation of the neural response to threat. *Psychological Science*, 17(12), 1032–1039.
- Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, 65, 333–371.
- Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S., & Gwaltney, J. M. (1997). Social ties and susceptibility to the common cold. *JAMA*, 277(24), 1940–1944.
- Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, 58, 10–23.
- Crocker, J., & Park, L. E. (2004). The costly pursuit of self-esteem. *Psychological Bulletin*, 130(3), 392–414.
- Crocker, J., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review*, 108(3), 593–623.
- Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. Jossey-Bass.
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the Experience-Sampling Method. *Journal of Nervous and Mental Disease*, 175(9), 526–536.
- Dane, E. (2010). Reconsidering the trade-off between expertise and flexibility: A cognitive entrenchment perspective. *Academy of Management Review*, 35(4), 579–603.
- Danckert, J., & Eastwood, J. D. (2020). *Out of My Skull: The Psychology of Boredom*. Harvard University Press.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18(1), 105–115.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668.
- Dietrich, A. (2003). Functional neuroanatomy of altered states of consciousness: The transient hypofrontality hypothesis. *Consciousness and Cognition*, 12(2), 231–256.
- Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared neural mechanisms of visual perception and imagery. *Trends in Cognitive Sciences*, 23(5), 423–434.
- Dimidjian, S., et al. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, 74(4), 658–670.
- Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind Over Machine*. Free Press.
- Duncker, K. (1945). On problem-solving. *Psychological Monographs*, 58(5).
- Eastwood, J. D., Frischen, A., Fenske, M. J., & Smilek, D. (2012). The unengaged mind: Defining boredom in terms of attention. *Perspectives on Psychological Science*, 7(5), 482–495.
- Farb, N. A. S., Segal, Z. V., et al. (2007). Attending to the present: Mindfulness meditation reveals distinct neural modes of self-reference. *Social Cognitive and Affective Neuroscience*, 2(4), 313–322.
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of Reinforcement*. Appleton-Century-Crofts.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58(2), 203–210.
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, 39, e229.
- Fiske, A. P. (1992). The four elementary forms of sociality. *Psychological Review*, 99(4), 689–723.

- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, 53(1), 25–44.
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *PNAS*, 102(27), 9673–9678.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
- Frost, R. O., & Hartl, T. L. (1996). A cognitive-behavioral model of compulsive hoarding. *Behaviour Research and Therapy*, 34(4), 341–350.
- Gazzaniga, M. S., & LeDoux, J. E. (1978). *The Integrated Mind*. Plenum.
- German, T. P., & Defeyter, M. A. (2000). Immunity to functional fixedness in young children. *Psychonomic Bulletin & Review*, 7(4), 707–712.
- Gibson, E. J. (1969). *Principles of Perceptual Learning and Development*. Appleton-Century-Crofts.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.
- Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies*, 29(1), 1–17.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Doubleday.
- Hackmann, A., Clark, D. M., & McManus, F. (2000). Recurrent images and early memories in social phobia. *Behaviour Research and Therapy*, 38(6), 601–610.
- Hagger, M. S., et al. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4), 546–573.
- Hayes, S. C., Wilson, K. G., Gifford, E. V., Follette, V. M., & Strosahl, K. (1996). Experiential avoidance and behavioral disorders. *Journal of Consulting and Clinical Psychology*, 64(6), 1152–1168.
- Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: A meta-analytic review. *PLoS Medicine*, 7(7), e1000316.
- Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality. *Perspectives on Psychological Science*, 10(2), 227–237.
- Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, 22(4), 337–349.
- Jacobson, N. S., et al. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, 64(2), 295–304.
- Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage*, 14(1), S103–S109.
- Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, 310(5745), 116–119.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, 98(6), 1325–1348.
- Kasser, T., & Ryan, R. M. (1993). A dark side of the American dream. *Journal of Personality and Social Psychology*, 65(2), 410–422.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, 28(2), 107–128.
- Kernis, M. H. (2003). Toward a conceptualization of optimal self-esteem. *Psychological Inquiry*, 14(1), 1–26.
- Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, 330(6006), 932.

- Koob, G. F., & Le Moal, M. (1997). Drug abuse: Hedonic homeostatic dysregulation. *Science*, 278(5335), 52–58.
- Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, 139(4), 665–682.
- Kross, E., et al. (2013). Facebook use predicts declines in subjective well-being in young adults. *PLoS ONE*, 8(8), e69841.
- Kross, E., et al. (2014). Self-talk as a regulatory mechanism: How you do it matters. *Journal of Personality and Social Psychology*, 106(2), 304–324.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 451–468.
- Leary, M. R. (2004). *The Curse of the Self*. Oxford University Press.
- Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology*, 68(3), 518–530.
- Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, 317(5841), 1096–1099.
- Limb, C. J., & Braun, A. R. (2008). Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS ONE*, 3(2), e1679.
- Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, 54(6).
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41(9), 954–969.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self. *Psychological Review*, 98(2), 224–253.
- Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133.
- Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*. MIT Press.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140.
- Neal, D. T., Wood, W., Wu, M., & Kurlander, D. (2011). The pull of the past: When do habits persist despite conflict with motives? *Personality and Social Psychology Bulletin*, 37(11), 1428–1437.
- Neff, K. D. (2003). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity*, 2(2), 85–101.
- Neisser, U. (1976). *Cognition and Reality*. W. H. Freeman.
- Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology*, 1(1), 35–59.
- Nesse, R. M., & Williams, G. C. (1994). *Why We Get Sick: The New Science of Darwinian Medicine*. Times Books.
- Nguyen, T. T., Ryan, R. M., & Deci, E. L. (2018). Solitude as an approach to affective self-regulation. *Personality and Social Psychology Bulletin*, 44(1), 92–106.
- Nigro, G., & Neisser, U. (1983). Point of view in personal memories. *Cognitive Psychology*, 15(4), 467–482.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.
- Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, 3(5), 400–424.
- Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, 3(2), 173–182.

- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116(2), 220–244.
- Prentice, D. A., & Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus. *Journal of Personality and Social Psychology*, 64(2), 243–256.
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9(2), 129–154.
- Raichle, M. E., et al. (2001). A default mode of brain function. *PNAS*, 98(2), 676–682.
- Reis, H. T., Clark, M. S., & Holmes, J. G. (2004). Perceived partner responsiveness as an organizing construct in the study of intimacy and closeness. In D. Mashek & A. Aron (Eds.), *Handbook of Closeness and Intimacy*. Erlbaum.
- Richards, D. A., et al. (2016). Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (COBRA). *The Lancet*, 388(10047), 871–880.
- Rohrer, D., Pashler, H., & Harris, C. R. (2015). Do subtle reminders of money change people's political views? *Journal of Experimental Psychology: General*, 144(4), e73–e85.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440–452.
- Schnall, S., Harber, K. D., Stefanucci, J. K., & Proffitt, D. R. (2008). Social support and the perception of geographical slant. *Journal of Experimental Social Psychology*, 44(5), 1246–1255.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories. *Cognitive Psychology*, 22(1), 36–71.
- Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, 15(7), 319–326.
- Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, 2, 395.
- Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings. *Journal of Applied Psychology*, 104(12), 1514–1534.
- Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering. *Annual Review of Psychology*, 66, 487–518.
- Solomon, R. L., & Corbit, J. D. (1974). An opponent-process theory of motivation. *Psychological Review*, 81(2), 119–145.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
- Swann, W. B., Pelham, B. W., & Krull, D. S. (1989). Agreeable fancy or disagreeable truth? Reconciling self-enhancement and self-verification. *Journal of Personality and Social Psychology*, 57(5), 782–791.
- Swann, W. B., Hixon, J. G., & De La Ronde, C. (1992). Embracing the bitter truth: Negative self-concepts and marital commitment. *Psychological Science*, 3(2), 118–121.
- Tackman, A. M., et al. (2019). Depression, negative emotionality, and self-referential language. *Journal of Personality and Social Psychology*, 116(5), 817–834.
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78(5), 853–870.
- Ulrich, M., Keller, J., Hoenig, K., Waller, C., & Grön, G. (2014). Neural correlates of experimentally induced flow experiences. *NeuroImage*, 86, 194–202.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. MIT Press.

- Verduyn, P., et al. (2015). Passive Facebook usage undermines affective well-being. *Journal of Experimental Psychology: General*, 144(2), 480–488.
- Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science*, 314(5802), 1154–1156.
- Wells, A., & Matthews, G. (1994). *Attention and Emotion: A Clinical Perspective*. Erlbaum.
- Wilson, T. D., et al. (2014). Just think: The challenges of the disengaged mind. *Science*, 345(6192), 75–77.
- Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, 114(4), 843–863.
- Wulf, G. (2013). Attentional focus and motor learning: A review of 15 years. *International Review of Sport and Exercise Psychology*, 6(1), 77–104.
- Wulf, G., Höß, M., & Prinz, W. (1998). Instructions for motor learning: Differential effects of internal versus external focus of attention. *Journal of Motor Behavior*, 30(2), 169–179.
- Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic Bulletin & Review*, 23(5), 1382–1414.