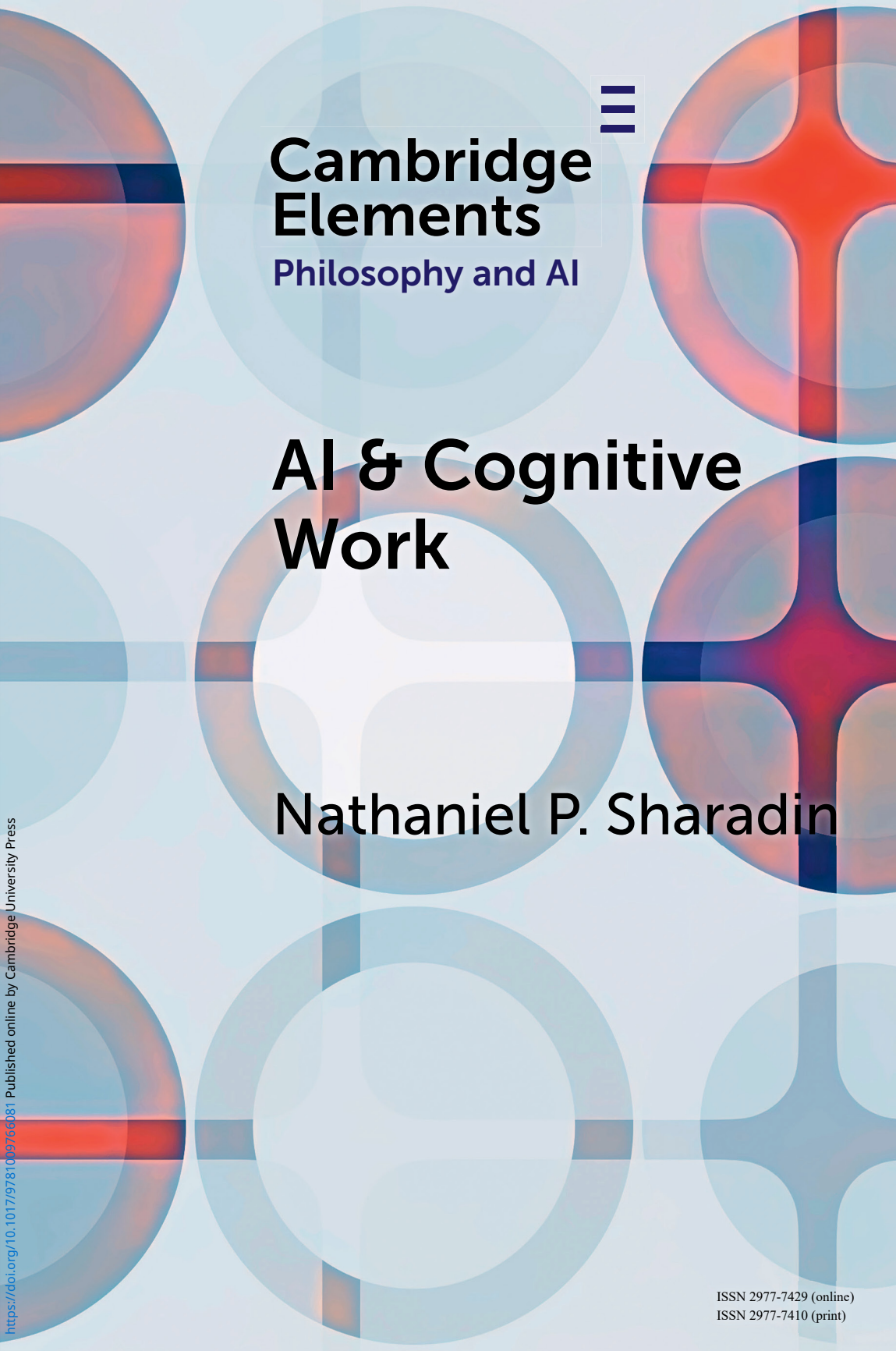


https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

ISSN 2977-7429 (online)
ISSN 2977-7410 (print)

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Cambridge Elements

Elements in Philosophy and AI
edited by
Herman Cappelen
University of Hong Kong

AI & COGNITIVE WORK

Nathaniel P. Sharadin
University of Hong Kong

CAMBRIDGE
UNIVERSITY PRESS

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press



Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781009766098
DOI: 10.1017/9781009766081

© Nathaniel P. Sharadin 2026

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, with the exception of the Creative Commons version the link for which is provided below, no reproduction of any part may take place without the written permission of Cambridge University Press & Assessment.

An online version of this work is published at doi.org/10.1017/9781009766081 under a Creative Commons Open Access license CC-BY-NC 4.0 which permits re-use, distribution and reproduction in any medium for non-commercial purposes providing appropriate credit to the original work is given and any changes made are indicated. To view a copy of this license visit https://creativecommons.org/licenses/by-nc/4.0

When citing this work, please include a reference to the DOI 10.1017/9781009766081

First published 2026

A catalogue record for this publication is available from the British Library

A Cataloging-in-Publication data record for this Element is available from the Library of Congress

ISBN 978-1-009-76609-8 Hardback
ISBN 978-1-009-76607-4 Paperback
ISSN 2977-7429 (online)
ISSN 2977-7410 (print)

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

For EU product safety concerns, contact us at Calle de José Abascal, 56, 1°, 28003 Madrid, Spain, or email eugpsr@cambridge.org.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

AI & Cognitive Work

Elements in Philosophy and AI

DOI: 10.1017/9781009766081
First published online: November 2026

Nathaniel P. Sharadin
University of Hong Kong

Author for correspondence: Nathaniel P. Sharadin, natesharadin@gmail.com

Abstract: Artificial systems now perform a striking range of cognitive tasks. This Element defends three claims. First, the status quo for cognitive work cannot survive. Whatever AI's trajectory, the package of features defining cognitive work – high skill requirements, credentials, wage premiums, resistance to automation, and low productivity growth – will be radically transformed. Second, the widespread displacement of humans from cognitive work, though it sounds alarming, may pose no special threat to human flourishing. Whatever makes work good for people is equally available outside cognitive work, and displacement enables leveraged gains in research and access to justice. Third, a world of ubiquitous cognitive assistance, though it sounds benign, carries distinctive moral risks, especially for children. Cognitive assistance works by reducing cognitive difficulty. But cognitive difficulty is necessary for the normal development of executive function – the capacities constitutive of autonomous agency. This title is also available as Open Access on Cambridge Core.

Keywords: future of work, cognitive development, AI and children, AI and jobs, human flourishing

© Nathaniel P. Sharadin 2026

ISBNs: 9781009766098 (HB), 9781009766074 (PB), 9781009766081 (OC)
ISSNs: 2977-7429 (online), 2977-7410 (print)

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Contents

1	Cognitive Work	1
2	Displacement, Assistance, Status Quo	11
3	Widespread Displacement	30
4	Ubiquitous Cognitive Assistance	44
5	Further Issues	65
	References	70

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

1 Cognitive Work

1.1 Introduction

Artificial systems can now perform a striking range of cognitive tasks. Whether their capabilities will continue to improve rapidly or plateau at current levels is uncertain. Either trajectory raises urgent questions about the future of cognitive work.

I defend three primary claims in this Element. First, the status quo for cognitive work is untenable. Second, the widespread displacement of humans from cognitive work, despite sounding alarming, may pose no special threat to human flourishing. Third, a world characterized by ubiquitous cognitive assistance, despite sounding benign, carries distinctive moral risks we've largely ignored. Together, these claims invert conventional wisdom.

In the remainder of this introduction, I do five things. I specify what I mean by "work" in this Element (§2). I characterize *cognitive* work and document its share of developed economies (§3). I identify five features of cognitive work that jointly form a distinctive package (§4). I sketch how recent advances in artificial intelligence (AI) threaten to disrupt this package (§5). I conclude with two assumptions that will guide the analysis in Section 2 (§6).

1.2 What Is Work?

What distinguishes work from everything else we do? There's no consensus answer. Historians (cf. Applebaum 1992) say one thing, sociologists (Caplow 1954) another, and philosophers, predictably, disagree with everyone, including each other (cf. Cholbi 2018). This is because what we call "work" depends partly on what question we're trying to answer (Furendal et al. 2024).

In this Element, I'm interested in the impact of recent advances in AI on work. I'll adopt an unapologetically *economic* conception of work: Work is the performance of tasks done with the expectation of remuneration (cf. Danaher 2017; Cholbi 2018). I adopt an economic conception because questions about AI's effect on labor markets seem especially fundamental in two respects.

First, the development of modern AI systems is primarily driven by the explicit goal of *automating* economic work. Among leading AI developers, OpenAI is explicit about their motivations: They aim to develop "highly autonomous systems that outperform humans at most economically

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

valuable work" (OpenAI 2024). An economic conception of work therefore tracks what the forces driving the development and deployment of this emerging technology actually target.

Second, and importantly, answers to noneconomic questions about AI and work often depend on answering economic questions. For examples: Whether people lose the social role of "worker" depends on displacement patterns, and whether inequality worsens depends on how productivity gains divide between capital and labor (Autor and Salomons 2018). Economics is, in many ways, upstream.

I make no claim to capturing everything we might call "work." An economic conception excludes unpaid domestic labor that's plainly "work" in an important sense. But it captures what's needed here, remains neutral about work's intrinsic value – it doesn't assume work is dignified or degrading, fulfilling or alienating – and aligns with how labor economists study technological change. I return to these analyses in Section 2.

1.3 The Rise of Cognitive Work

What makes work specifically *cognitive*? Cognitive work, as opposed to *not*-cognitive work, primarily involves mental, as opposed to physical, activity (Autor et al. 2003). A useful heuristic: If a job is mostly done at a desk or computer, it's likely cognitive work. Lawyers, accountants, writers, analysts, programmers, managers, academics, teachers, and consultants are all paradigm cases of cognitive workers. Carpenters, truck drivers, nurses, cooks, and machinists are paradigm cases of *not*-cognitive workers, though many of these jobs involve significant cognitive components. I return to the mixed nature of jobs in Section 3.

We can further distinguish between *routine* and *nonroutine* cognitive work (Autor 2015). Routine cognitive work tends to follow explicit procedures. Nonroutine cognitive work, by contrast, requires judgment and contextual sensitivity. Paradigmatic examples of routine cognitive work include: processing invoices, checking forms for completeness, and basic data entry. Paradigmatic examples of *non*routine cognitive work include: legal reasoning, diagnosis, strategic planning, and persuasive writing. As we'll see later in this section (and again in Section 2), what's distinctive about modern AI systems is their potential for automating *non*routine cognitive work.

The central empirical fact about cognitive work is that it comprises a large and increasing share of developed economies. Preindustrial economies were dominated by physical labor, most of it agricultural. For

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

instance, in the United States, agriculture employed over 40% of workers in 1900 (Autor 2014). Industrialization shifted workers into manufacturing – still largely physical, but increasingly routinized. The postindustrial transition moved labor again, this time into services and what we now recognize as (both routine and nonroutine) cognitive work. Manufacturing employment peaked in the mid-twentieth century at around 25% of U.S. workers and then declined to roughly 10% by 2010 (Autor 2015). By 2010, occupations characterized by cognitive work comprised over 40% of U.S. employment, whereas agricultural jobs comprised less than 2% (Goos et al. 2014; Autor 2014).

The relationship between educational attainment and jobs tells a similar story: A steadily increasing proportion of jobs require specialized training (Autor 2014), and returns to education have risen steadily for the past century (Carneiro et al. 2011). The same pattern appears across wages: cognitive workers command a steadily increasing wage premium (Autor and Handel 2013), reflecting both the investment required to develop cognitive skills and employer demand for them (Goldin and Katz 2007).

What's driving this trend? Three dynamics stand out.

First, the increasing automation of routine physical tasks pushed labor toward roles that machines couldn't perform. Until recently, this meant pushing workers toward (routine and nonroutine) cognitive work (Acemoglu and Restrepo 2018).

Second, and simultaneously, globalization off-shored much manufacturing to countries where labor costs were lower (Autor et al. 2016). Meanwhile, cognitive work proved harder (though not impossible) to offshore (Blinder and Krueger 2013; Blinder 2006).

Third, rising incomes shifted demand toward services. As societies grow wealthier, households spend proportionally more on healthcare, education, and other services – Engel's law (1857). These sectors are disproportionately cognitive intensive. This demand-side shift complemented the supply-side pressures from automation and off-shoring, pulling labor into cognitive work as other forces pushed it there.

The overall picture looks significantly different in developing economies, where manufacturing employment may still be rising (Kruse et al. 2022). In this Element, I focus on developed economies (e.g., the United States) where cognitive work's share of the economy is largest. If AI meaningfully transforms cognitive work, its impact will be greatest where cognitive work's share is greatest. I return briefly to questions about how things might be different in developing economies in Section 5.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

1.4 Five Features of Cognitive Work

Five features form cognitive work's economically distinctive core.

Feature 1: *High skill requirements*. Cognitive work typically demands complex skills. These skills take years (sometimes decades) for humans to develop, often requiring extended practice (Ericsson et al. 1993). Much of the knowledge involved in exercising such skills is tacit in the twofold sense that it is difficult to articulate or codify and that it is (therefore) often learned by doing rather than by explicit instruction (Polanyi 1966). In this way, cognitive work contrasts with work where procedures can be standardized and skills acquired relatively quickly (Autor et al. 2003).

Feature 2: *Credentials*. Because the skills required for cognitive work are challenging to observe directly, employers rely on credentials as useful proxies for their presence. For this reason, most cognitive work requires formal qualifications, such as educational degrees. Education serves partly as genuine preparation for the performance of specific cognitive tasks and partly as a *signal* that the worker can handle a wide variety of cognitively demanding tasks (Spence 1978; Becker 1994).

Feature 3: *High wages*. Cognitive workers earn significantly more, on average, than not-cognitive workers (Autor et al. 2008). This wage premium is the labor market's natural reflection of the first two features defining cognitive work: scarcity of cognitive skills and the investment required to acquire credentials (cf. Becker 1994). The premium isn't uniform. There is significant variation within cognitive work, and some not-cognitive work commands high wages. But the broad pattern holds (Autor and Handel 2013).

Feature 4: *Historical resistance to automation*. Earlier rounds of automation targeted physical and routine (cognitive and not-cognitive) tasks. But (especially nonroutine) cognitive work has remained stubbornly human. This is sometimes called the Moravec paradox: tasks that seem easy for humans (recognizing faces, understanding context) have proved hard for machines, while tasks that seem hard for humans (rapid calculation, searching large databases) have proved easy for machines (Moravec 1988). Cognitive work, especially nonroutine cognitive work, has remained squarely in the "hard for machines" category for the past several decades (cf. Brynjolfsson and McAfee 2014).

Feature 5: *Limited productivity growth*. Manufacturing saw dramatic productivity gains over the twentieth century: output per worker increased many times over (Gordon 2016). Productivity in cognitive work, meanwhile, has largely stagnated. One central reason is that cognitive output

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

depends on the time and attention of skilled humans. This infects sectors that depend on cognitive work with what economist William Baumol called a "cost disease" (Baumol 1967, 2012). Wages in low-productivity-growth sectors must rise to compete for workers with high-productivity-growth sectors. But if productivity isn't rising to match, the result is that goods that depend on cognitive labor (e.g., education, healthcare, legal services) become relatively more expensive over time as a result of differential productivity growth (Nordhaus 2006).

These five features are not independent; instead, they form a mutually reinforcing package. High skill requirements necessitate extensive education and credentialing, which in turn mean that cognitive workers can demand high wages. These high wages make cognitive work an attractive target for automation. But the complexity and context-sensitivity of (especially nonroutine) cognitive work has historically made automation stubbornly difficult in practice. This resistance to automation in turn keeps productivity growth low, which, given the relative scarcity of skilled human labor, drives up relative costs.

This package of features helps explain why cognitive workers have enjoyed a privileged position in developed economies over the past half-century. It also explains why AI's impact on cognitive work could be so consequential: modern AI systems threaten to dismantle the entire package.

1.5 AI and the Transformation of Cognitive Work

Modern AI systems designed to automate cognitive work are built on large language models (LLMs) – artificial neural networks trained on vast quantities of text to predict and generate language. Pre-training works by exposing a model to massive corpora and adjusting the model's (now *trillions* of) parameters to improve accuracy at next-word (token) prediction. Post-training shapes a model's outputs, making responses more helpful, accurate, and aligned with user intent through techniques such as reinforcement learning from human feedback (Christiano et al. 2017). Recent *reasoning models* extend this paradigm further, training systems to work through responses step-by-step (Wei et al. 2022) rather than producing immediate responses, and rewarding models for showing their work and penalizing shortcuts (DeepSeek-AI et al. 2025). Developers then add robust software scaffolding ("harnesses"; Anthropic 2025) that enable systems to use tools, retrieve information, and pursue goals across multiple steps (Yao et al. 2023).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

The resulting systems exhibit broad competence across a variety of cognitive tasks that completely stymied earlier AI systems (cf. Harding and Sharadin 2024). Multimodal training extends these capabilities to images, audio, and video (Gemini Team et al. 2024). I return to an overview of current capabilities in Section 2.

These systems have significant limitations. Perhaps most famously, they confidently generate plausible-sounding falsehoods, so-called hallucinations, though rates of hallucination are falling rapidly (Huang et al. 2025). Still, these systems do exhibit weird failure modes (cf. Bommasani et al. 2022), and they are not "generally," let alone "super," intelligent (cf. Bostrom 2014). Whether their capabilities will continue to improve rapidly or plateau is uncertain. But even existing capabilities are economically significant, and the trajectory points toward systems performing an ever-wider range of both routine and nonroutine cognitive tasks over increasingly long time-horizons (Acemoglu and Restrepo 2020a; Kwa et al. 2025).

This matters because earlier technologies targeted either physical or at most *routine* cognitive tasks for automation. By contrast, modern AI systems directly target *nonroutine* cognitive tasks: the core of most (high-skill, credentialed, high-pay) cognitive work. Consider, briefly, how AI capable of performing high-skill, nonroutine cognitive work might affect each of our five features.

Feature 1: *High skill requirements*. AI may lower the skill floor for reliably performing many cognitive tasks. Work that previously required years of training may become accessible to relative novices equipped with capable AI assistants. This would democratize access to cognitive work by lowering skill requirements.

Feature 2: *Credentials*. If AI can reliably perform a cognitive task, then a credential that signals expertise in that task will lose value, or even become obsolete. Educational institutions adapt slowly; labor markets may move faster. The result could be reduced barriers to entry for cognitive work, new forms of credentialing based around AI, or both.

Feature 3: *High wages*. Cognitive workers earn more than not-cognitive workers because the skills required to perform cognitive tasks are relatively scarce. This scarcity creates pricing power for cognitive workers selling their labor into the market. But AI threatens this dynamic directly. If AI systems can match (or exceed) human performance on economically valuable cognitive tasks, the effective supply of labor available for performing these tasks in the market radically expands. The wage premium

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

cognitive workers have enjoyed may therefore erode as AI makes an input into cognitive firms' production functions (much) more abundant.

Feature 4: *Historical resistance to automation*. Nonroutine cognitive tasks require flexibility, judgment, and the ability to handle novel situations. Earlier, rule-based, "Good Old-Fashioned" AI systems couldn't provide this flexibility (cf. Boden 2014). Recent AI systems, as we've noted, are different in kind. They exhibit broad competence across many nonroutine cognitive tasks and can increasingly flexibly handle novel contexts beyond their training. The philosophical question of whether these systems exhibit genuine "understanding" (cf. Cappelen and Dever 2025; Bender et al. 2021) is orthogonal to their economic impact: They can already perform cognitive tasks that are economically valuable. The barrier that protected nonroutine cognitive work from automation is already eroding.

Feature 5: *Limited productivity growth*. As discussed, Baumol's cost disease constrains productivity in cognitive work because output depends on scarce human attention. AI may break this constraint: producing additional cognitive capacity currently requires training humans over years; AI systems can be copied at low marginal cost and deployed at scale.

1.6 Plausible Assumptions about Capabilities

Given these possibilities, what kind of future for cognitive work should we actually expect? The most significant source of uncertainty concerns the trajectory of AI systems' capabilities. AI capabilities might continue improving rapidly, or even speed up – potentially *a lot* (cf. Bostrom 2014; Davidson et al. 2025; but see Thorstad 2024); they could plateau soon or follow some more unpredictable, "jagged" trajectory (cf. Gans 2026).

In this Element, I'll aim to bracket (at least some of) this deep empirical uncertainty by examining what follows from assumptions at opposing ends of the *plausible range* of views about how capabilities will change over time (Figure 1). This is a standard philosophical strategy (cf. Hajek 2016).

What makes an assumption plausible? In this context, I aim for *conservatism*: scenarios that involve minimal departures from how the world actually is (cf. Lewis 1973). Applied to AI capabilities, the question becomes: Which future trajectory requires the least deviation from established trends?

One assumption is that *current directions and rates of trends continue*: AI capabilities keep improving (the direction of the trend) at roughly the pace (the rate of the trend) of recent years. This assumption holds fixed the

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

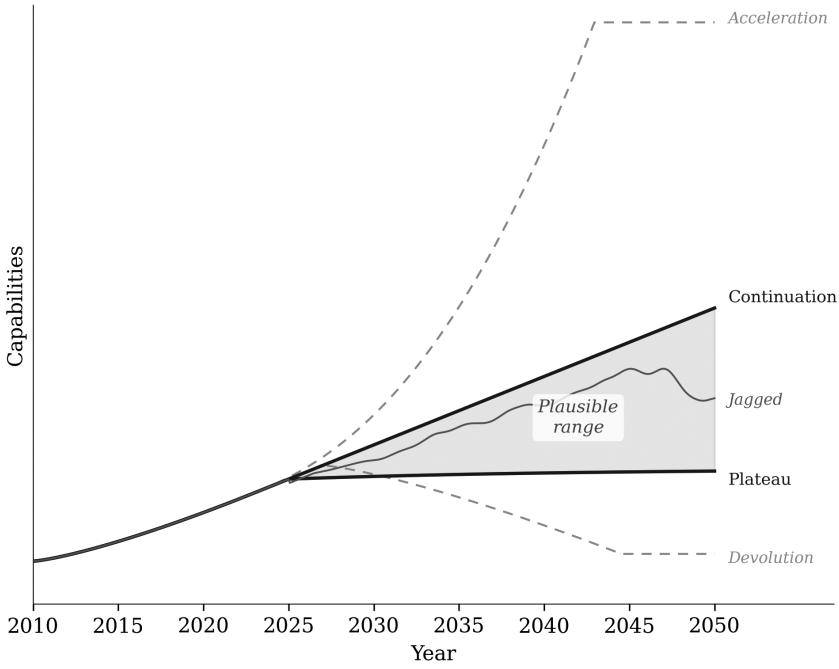


Figure 1 Stylized representation of plausible trajectories for AI capabilities. Solid lines represent the two assumptions analyzed in this Element: *continuation* (capabilities improve at roughly recent rates) and *plateau* (capabilities flatten at roughly current levels). The shaded region marks the plausible range. Dashed lines represent trajectories set aside: acceleration and devolution. The jagged trajectory illustrates uneven progress within the plausible range.

research paradigm (Vaswani et al. 2017), the scaling relationships (Kaplan et al. 2020), the investment patterns (Stanford HAI 2025), and the cadence of progress in terms of observed capabilities that has been the result of these forces (Gemini Team et al. 2024; OpenAI 2022; Anthropic 2024). Call this the *continuation* assumption.

One might object that the appearance of conservativeness in continuation is misleading; perhaps a truly conservative set of assumptions would anchor to the current state of technology, rather than locking in a rate of change that (as I'll argue in Section 2) guarantees radical disruption. But this confuses two questions. The first is whether, if things continue roughly as they are on track to do at present, the medium-term future will be radically different from the present in surprising ways. The second is whether futures where things are radically different from the present are similar to the present. Obviously, the answer to the second question is "No." But this

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

is irrelevant to answering the first question. Continuation might lead to a radically different world. But that doesn't make "things will go on roughly as before" an exotic assumption.

There are two salient alternatives to current trends continuing apace. One is that trends accelerate. Call this the *accelerationist* assumption. I will ignore this possibility in what follows, for three reasons. First, it complicates the analysis considerably. Second, the complications it introduces are not material to my substantive normative claims. (Though: I note some provisos throughout.) Third, I think everyone should agree that, were current trends in AI capabilities to accelerate, the world would rapidly become sufficiently weird as to make it challenging to say anything precise about it in advance.

The other salient set of assumptions involves trends slowing down or even flattening out entirely. Since I'm aiming to analyze possibilities at the extreme ends of plausibility, I'll take the *most* conservative of these options, which is flattening out. Assume trends plateau, and we get approximately no new improvements in AI capabilities. Call this the *plateau* assumption.

You might think there's a third set of plausible assumptions that yield the result that AI systems' capabilities actually decline. But assumptions required to yield a devolutionary result make little sense. What could make it true that AI systems get *worse* at performing cognitive tasks they *already* perform at some given skill level? We'd have to imagine a global smashing of the looms, which I do not think is a plausible assumption. I return to this issue briefly in Section 4.

So, setting aside acceleration and devolution, these two sets of assumptions – continuation and plateau – bracket the plausible range (Figure 1). Continuation assumes recent upward trends persist; plateau assumes they flatten but still respects the one-directional ratchet of technological capability.

Consider, first, continuation: AI systems' capabilities continue improving at something like the recent (very rapid) rates. Intuitively, in this world we will (quite soon) develop AI systems that succeed at OpenAI's stated mission: We will inhabit a world with systems that can perform (most) economically valuable cognitive work at roughly human level. In this case, I'll argue we likely face the widespread *displacement* of humans from cognitive work.

Now consider plateau: AI systems' capabilities stall roughly where they are now. Intuitively, in this world we may get meaningful improvements to the usefulness of existing systems for cognitive work through diffusion and

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

falling costs. But, for whatever reason, the current research paradigm "hits a wall" (cf. Marcus 2022), and AI turns out to be "normal technology" (Narayanan and Kapoor 2025). In this case, I'll argue we face a future of ubiquitous cognitive *assistance* that transforms cognitive work even while human workers remain in the loop.

It's tempting to think that the most likely outcome, especially in the medium term, is somewhere in the middle–a mixture of displacement in some domains and assistance in others. This intuition follows from standard diffusion dynamics. New technologies typically spread unevenly, and different sectors integrate them at different rates (Geroski 2000). But AI may not follow normal diffusion patterns, for at least two reasons.

First, AI differs structurally from earlier technologies in important ways. While sectorspecific customization is still required, the same AI system (more carefully, the same foundation model; cf. Bommasani et al. 2022) can often be adapted across domains with relatively modest effort. Intuitively, a system that learns legal reasoning can learn medical diagnosis, too. And once deployed, AI systems scale near-instantly at near-zero marginal cost. (I say more about the physical constraints on AI deployment in Section 2.)

Second, the diffusion of AI will be shaped by competitive pressures even more intense than in previous technological transitions. Given the massive opportunity for capturing productivity gains (labor's share of GDP is around 50%, and cognitive labor is almost half of that), extreme competitive pressure between firms may compress diffusion timelines. And setting aside firms, the case of AI is attended by an increasingly intense economic and strategic competition between "great powers" such as the US and China and, to a lesser extent, Europe (Ding 2024). The mixed case may therefore be more likely to be a transition state rather than a stable equilibrium.

I aim to sidestep these extremely delicate empirical questions. My hope is that analyzing unmixed cases under assumptions at extreme ends of the plausible range can give us tools to think about various combinations thereof. And importantly, as part of this analysis I'll argue that there are *no* plausible assumptions that preserve the status quo for cognitive work: Whatever happens, we're headed for a world where the package of features that has defined cognitive work for the last several decades will be radically transformed.

The following sections take up these arguments in more detail. Section 2 develops the argument just sketched: that conservative assumptions support widespread displace- ment, that *extremely* conservative assumptions

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

still yield ubiquitous assistance, and that only wishful thinking supports the status quo. Sections 3 and 4 turn to normative questions about these outcomes. Section 3 argues that widespread displacement poses no special threat to human flourishing. Section 4 argues that ubiquitous assistance carries distinctive moral risks we've largely ignored – risks that extend beyond labor markets to children's developmental environments. Section 5 broadens the lens to questions this Element lacks space to address in depth.

2 Displacement, Assistance, Status Quo

2.1 Introduction

Section 1 established that cognitive work is central to developed economies and displays five distinctive features that form a mutually reinforcing package. This section asks: What happens to this package under plausible assumptions about AI's trajectory? As I argued in Section 1, there are two assumptions about AI capabilities that are especially salient.

If capabilities continue improving, we face *widespread displacement*. This is the intuitively dystopic "drop-in remote worker" scenario that is the explicit goal of firms developing and deploying this technology (cf. OpenAI 2024; Google, 2025).

If capabilities instead plateau, we face *ubiquitous assistance*. This is the intuitively utopic "co-intelligence" (Mollick 2024) or "extended mind" (cf. Chalmers and Clark 1998) scenario optimists say we should aim for.

Maybe you think there's a third possibility: the *status quo*. A status quo future for cognitive work is one where AI becomes a useful productivity tool while leaving the distinctive economic structure of cognitive work intact. Consider the spreadsheet: It transformed how accountants work by automating calculation and enabling complex modeling, but it didn't, for example, erode wage premiums or eliminate credential requirements. Status quo assumes AI follows roughly this pattern.

I'll argue that status quo isn't compatible with either of our plausible assumptions. Status quo would require an implausible quinella of assumptions: AI progress halts and already existing systems don't diffuse.

2.2 The Task-Based Framework

What determines whether (and how) AI disrupts the package of features that defines cognitive work? To answer that, we need a framework for how technology affects labor markets.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

One natural approach is to think in terms of whether a "job" can be automated. But this approach obscures more than it reveals. A lawyer doesn't "lawyer." A lawyer drafts documents, counsels clients, negotiates deals, reads briefs, appears in court, writes emails, manages junior associates, takes clients out to lunch, and so on. Some of these tasks may be automatable; others may not. Job-level analysis misses this variation. Worse, it treats jobs as fixed targets. But when only some tasks in a given job are (liable to being) automated, those jobs don't simply vanish; instead, they reconfigure around the remaining tasks, and new tasks often emerge.

The solution is to switch from a *job*- to a *task*-based framework: Jobs are bundles of heterogeneous tasks, so tasks are the relevant unit of constant analysis.[1] A task is a discrete unit of productive activity.[2] A job bundles multiple tasks assigned to a worker. An occupation is a category of similar jobs. Labor markets allocate workers to tasks, mediated by jobs and occupations. Automation, meanwhile, operates at the *task* (rather than job or occupational) level. When a machine can perform a task more cheaply or reliably than a human, substitution pressure emerges, and the task migrates from human labor to capital (Autor and Handel 2013).

This framework thus identifies two countervailing forces that jointly determine the effects of emerging technology on labor markets: *displacement* and *reinstatement* (Acemoglu and Restrepo 2019). Displacement effects are straightforward: When the cost of automated task performance at a given level of reliability falls below the cost of human labor, demand for human labor in that task falls. And if a job consists mostly of such tasks, the occupation will contract, and the job itself may disappear. If only some of a job's tasks meet this threshold, then the job shrinks or reconfigures around the tasks that remain.

Reinstatement effects push in the opposite direction. Technological change creates new tasks that didn't previously exist. The specific content of these tasks is often hard to predict in advance – for instance, "cloud architect" would have been an unintelligible job description thirty years ago. But the pattern is relatively robust (Autor 2015).

[1] Key statements of the framework include (Autor et al. 2003; Autor and Salomons 2018; Acemoglu and Restrepo 2017, 2019, 2020b; Autor and Handel 2013; Acemoglu and Autor 2011).

[2] The framework's decomposition of jobs into tasks already embeds normative choices – about where to draw task boundaries and who decides which tasks are candidates for automation. I'll work within these choices while noting that they aren't value-neutral.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Daron Acemoglu and Paul Restrepo characterize this overall dynamic as a "race" between displacement and reinstatement (Acemoglu and Restrepo 2018). When new task creation keeps pace with automation, employment and wages can grow even as individual tasks are automated away (i.e., are reassigned from labor to capital). Workers displaced from automated tasks can then move into newly created tasks. Displaced workers don't always make these transitions themselves; often it is the next generation that enters a transformed labor market. But, at least in the past, new categories of work have always emerged in developed economies to absorb labor in the aggregate: there isn't a fixed amount (a "lump") of work to be done.

The framework's central question about the relationship between AI and cognitive work, then, is whether this "race" between displacement and reinstatement continues apace or whether one side pulls decisively ahead. If AI automates existing cognitive tasks faster than new tasks emerge, or if AI can perform new cognitive tasks as readily as old ones, then displacement dominates within cognitive work. If new cognitive tasks continue to emerge and, crucially, humans retain advantage in them, then reinstatement keeps pace.

Before applying the framework, we need to know what existing AI systems can actually do.

2.3 The Current State of AI for Cognitive Work

Since I'm analyzing outcomes under different capability assumptions rather than forecasting which capabilities are likely to emerge, we can set aside technical details.[3] What matters is establishing current capabilities as a baseline. Trends continuing apace project forward from this baseline, and trends plateauing assume we stay roughly where we are.

So, what can these systems already do in domains that matter for cognitive work? Early estimates suggested nearly half of U.S. employment faced high risk of computerization, though this predated LLMs and focused on routine cognitive tasks (Frey and Osborne 2017). More recent AI-specific measures converge on a different finding: high-skill cognitive work faces disproportionate exposure to existing AI capabilities.

The pattern appears consistently across different approaches to measurement. Continuous exposure scores mapping AI capabilities to occupational task requirements show that high-skill, high-wage occupations

[3] For an excellent conceptual overview of the core underlying architecture of modern LLMs (Vaswani et al. 2017), see Levinstein (2023).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

in cognitive sectors are particularly highly exposed (Felten et al. 2018). Analysis of AI-patent text and job task descriptions for highly educated workers shows substantial overlap (Webb 2019). Eloundou et al. (2023) assessed exposure to LLMs specifically, estimating that approximately 80% of the U.S. workforce has at least 10% of their tasks exposed to LLMs, with exposure rising very sharply at higher wage levels.

These exposure patterns hold globally and have expanded along with recent improvements in AI capabilities. Cognitive and analytical tasks show high exposure across countries, with significant variation based on industrial and occupational structure (Gmyrek et al. 2023). A recent analysis of both advanced and emerging economies finds that nearly 40% of global employment faces high exposure to AI, with advanced economies showing even higher rates due to their concentration in cognitive-intensive sectors (Pizzinelli et al. 2023). AI exposure has accelerated dramatically in recent years, particularly for professional and technical occupations (Felten et al. 2021, 2023). Cognitive work faces substantial exposure to existing AI capabilities, even before accounting for further improvements in the technology or its integration into workflows.

2.4 The Simple Case for Displacement

The task-based framework identifies two countervailing forces: displacement and reinstatement. The question is whether one pulls decisively ahead. Suppose AI capabilities keep improving at roughly recent rates, systems steadily become more reliable and general, and costs continue to fall.

Then, the widespread displacement of humans from cognitive work appears likely to dominate the race, since AI systems will handle new cognitive tasks as quickly as they emerge. Meanwhile, falling costs will intensify the economic pressure to substitute AI for human labor year by year. For the race to end differently – that is, for reinstatement to keep pace – something would have to change from current trajectories.

Maybe you think this is too quick. As I emphasized earlier, technological change historically creates new tasks even as it automates old ones. But for new task creation to preserve the distinctive character of cognitive work specifically, such new tasks must satisfy two conditions: They must be *cognitive tasks* (i.e., not care work, not physical services, not entertainment), and, crucially, they must be tasks humans can perform better or more cheaply than AI systems. Given our working assumption, it beggars belief that new tasks will meet either of these conditions.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

In 2016, the ancestors of present-day frontier systems couldn't write a coherent sentence. Today, they demonstrate reasoning capabilities that rival specialized professionals (Phan et al. 2025) and, as we discussed earlier, can already complete a range of economically valuable cognitive tasks to very high degrees of reliability. If we assume this trend continues, there is little reason to believe future versions of these systems will be incapable of performing whatever new cognitive tasks emerge, likely at very low costs.

2.5 Resisting Displacement: Stalled Improvement

There are two ways to resist this conclusion. The first denies our assumption that current trends continue at roughly current rates. The second grants our assumption but argues that equilibrium effects nevertheless preserve substantial human roles in cognitive work.

Perhaps scaling hits diminishing returns (Hoffmann et al. 2022), or we're approaching the limits of current techniques (Marcus 2022), or certain capabilities require architectural breakthroughs we haven't discovered (Chollet 2019; Lake et al. 2017). Perhaps some limitations reflect something deep about what statistical learning from text can achieve (Browning and LeCun 2023). Perhaps available training data is nearly exhausted (Villalobos et al. 2022). Perhaps compute constraints become binding as energy and infrastructure costs escalate (Sevilla and Roldán 2024). Perhaps governance (in the name of workers, safety, ethics, democracy, or whatever) intervenes to deliberately slow progress in capabilities (Bullock et al. 2023).

Any or all of these may prove true! But notice what this objection does in all its forms: It says that trends in capabilities will stall, and it thereby shifts us toward the plateau assumption. I take up this assumption below.

2.6 Resisting Displacement: Equilibrium Effects

A different strategy for resisting displacement grants that AI capabilities steadily improve and that costs fall (per our working assumption) but argues that various *equilibrium effects* nevertheless preserve a substantial role for humans in cognitive work. I'll consider five.

2.6.1 Demand Expansion

First, consider *demand expansion*. Automation in a sector increases productivity, which lowers prices and can expand demand for the sector's output (Autor 2022). In some cases, demand increases can actually outpace

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

supply increases–the "Jevons effect" (Alcott 2005). In the present case, cheaper cognition might expand what's worth cognizing.

This mechanism is real. But on the assumption that AI capabilities keep improving, demand expansion likely can't keep pace. Here's why: As AI gets better at cognitive tasks, productivity gains compound. For demand expansion to prevent displacement, demand would need to grow at least as fast. But there's no reason to think demand for legal analysis or medical diagnosis expands without limit simply because it gets cheaper. Instead, at some point, we've presumably done all the legal analysis worth doing. Put in economic terms: demand expansion absorbs productivity gains only if price elasticity of demand persistently exceeds one. But demand curves for cognitive services aren't infinitely elastic! And at compounding rates, they don't need to be steep for long before reaching saturation.

2.6.2 Task Complementarity

Second, consider *task complementarity*. Automation of some tasks can increase the value of complementary tasks (Milgrom and Roberts 1990). Everyone's favorite example is ATMs: When ATMs automated cash handling, they made bank branches cheaper to operate. Banks responded by opening more branches and hiring more workers to handle loan consultations and complex account issues. Similarly, if AI handles a subset of cognitive tasks, human judgment on the remaining cognitive tasks could become more valuable.

There are two versions of this idea. The weaker version holds that AI capabilities will plateau at routine cognitive tasks, while complex judgment remains durably human. This would create a stable refuge. But AI that improves on existing trends year over year is unlikely to remain confined to routine cognitive work.

The stronger version says that even as AI capabilities expand, the remaining cognitive tasks assigned to humans become more valuable precisely because they're complements to an ever-growing AI-performed portion. This is correct as far as it goes: Complemen-tarity can technically raise the marginal product of the remaining human contributions *indefinitely* even under continuation. But indefinitely growing marginal product per worker is perfectly compatible with sharply reduced demand for workers overall. Indeed, it is what we would expect. This possibility is not theoretical. It is the story of American manufacturing: output up, productivity up, and employment down by *two-thirds*. Indefinite complementarity might therefore create a smaller corps of highly valuable human cognitive workers, but this wouldn't preserve cognitive work at scale. As I'll argue

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

below regarding O-ring effects, mechanisms that make remaining workers more productive in general can't explain why existing employment levels in cognitive work would persist.

2.6.3 Wealth Effects

Third, consider *wealth effects*. Automation generates wealth through productivity gains, and wealth generates demand for goods and services, including labor-intensive ones (Autor 2022). If AI dramatically increases productivity in cognitive work, the resulting wealth could fund employment in sectors far removed from the automated tasks themselves.

I expect economists are correct that a world where all cognitive work is automated is a much – maybe *much much* – richer world. But setting aside distributional questions, wealth effects are perfectly compatible with the collapse of cognitive work as a distinctive category. For instance, AI-driven productivity gains might create wealth that funds more massage therapists and home health aides. But that would not be the preservation of cognitive work as we know it! It would be the widespread displacement of human beings from cognitive work into other kinds of work. If anything, wealth effects facilitate displacement by ensuring displaced cognitive workers can find other employment – just not in cognitive work.

2.6.4 Comparative Advantage

Fourth, consider *comparative advantage*. Comparative advantage says that what matters for task allocation isn't who performs a task best in absolute terms but who is *relatively* better (cf. Buchanan 1991). Intuitively, if a senior lawyer is 3× faster than a junior at drafting but only 2× faster at client meetings, it is relatively costly for the senior to spend time in meetings (since they could be drafting instead). Substitute AI for the senior lawyer, and this might suggest humans might retain whatever cognitive tasks AI is relatively less dominant at.

Comparative advantage only operates under constraints imposed by scarcity. What would make the capacity of AI systems to accomplish cognitive tasks scarce? A single AI system can't do task A and task B simultaneously. But this trivial bottleneck only helps humans if scaling or multitasking is constrained. To make the argument for displacement as difficult as possible, suppose there are persistent architectural constraints on multitasking.

That leaves constraints on firms' ability to deploy more than one AI system. What could impose such constraints? The most obvious candidates

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

are scarcity in both compute and electricity. But these constraints are rapidly loosening, not tightening: data centers are being built at unprecedented scale, power infrastructure is expanding, and efficiency gains in algorithms mean that each unit of compute accomplishes more.[4]

But suppose scaling does face *some* binding constraints in the form of relatively scarce compute and electricity. Even so, it seems to me that the alleged refuge offered by comparative advantage to human cognitive workers will be cold comfort. There are two reasons for this.

First, for comparative advantage to preserve human roles in cognitive work, humans' relative disadvantage must be smallest in at least some cognitive tasks. But why think this is likely to be true? These systems are being *deliberately designed* to efficiently accomplish *cognitive tasks in particular*. If anything, a very different outcome seems much more likely, given the capabilities trajectory we're assuming here: humans retain comparative advantage at various *not*-cognitive tasks (robotics is hard) even as their disadvantage compared to AI systems continues to grow for cognitive tasks in particular. Worse, even if humans retain comparative advantage with respect to *some* cognitive tasks, why think this much smaller subset of cognitive tasks will be large enough to support the status quo for cognitive work?

But suppose (what seems incredible) that human cognitive workers retain comparative advantage with respect to *most* cognitive tasks. Still, this tells us nothing specific about the wage those workers will demand in an unrestricted market for their labor. That wage will be capped by the opportunity cost of deploying an AI system instead. But as AI gets increasingly cheaper and increasingly abundant, that opportunity cost drops. Correspondingly, the ceiling on human cognitive workers' wages will predictably drop, potentially far below subsistence (Korinek and Suh 2024).

2.6.5 O-Ring Automation

Finally, consider a recent proposal involving concentration effects (Gans and Goldfarb 2026). Gans and Goldfarb show that when production requires multiple tasks whose qualities multiply – an "O-ring" production

[4] Contrary to some reporting, water does not appear to be a significant constraint on AI scaling. Claims about AI's water consumption often conflate cooling water circulating ("used") in closed loops with water actually "consumed," or simply misreport the numbers. For example, Karen Hao's (2025) widely cited reporting on this topic makes both mistakes. Hao's reporting in her case study was factually incorrect – off by *three orders of magnitude* (Masley 2025) – and it misleadingly claimed that data centers "consumed" the water they use, when in fact it was returned to source (or evaporated) after being used for cooling. I briefly consider artificial constraints on scaling, such as regulation, below.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

function, following Kremer (1993) – automating some tasks can raise the marginal product of remaining human labor. Intuitively: if a worker allocates fixed time across tasks, automating only *some* tasks lets the worker concentrate remaining time on fewer tasks, raising quality on each. Since output is multiplicative in task qualities, this concentration can scale the value of human work in remaining bottleneck tasks. But given our assumption that AI capabilities keep improving, this mechanism fails to preserve the status quo for cognitive work for two reasons.

First, the O-ring mechanism operates at the level of individual workers, not labor markets. The model therefore explains why a certain kind of concentration in production functions makes each employed worker more valuable. But it tells us nothing about how many such workers an economy needs. As we noted earlier, indefinitely growing marginal product per worker is compatible with sharply reduced demand for workers overall.

Second, Gans and Goldfarb's model assumes that task quality equals labor productivity times hours, with no diminishing returns (Gans and Goldfarb 2026, 3). This is a bizarre assumption if we are trying to model cognitive work in particular. Their model works if we relax this assumption to allow for concave quality functions (i.e., we allow diminishing returns in quality for time invested). But this doesn't resolve the underlying problem, which is that, intuitively, there's only so much extra human attention you can give a cognitive task before you hit a hard quality ceiling.

More carefully: Gans and Goldfarb's model shows that when workers are displaced from some (cognitive) tasks, the remaining tasks are "performed with disproportionately high care and are, therefore, disproportionately hard to replace" (Gans and Goldfarb 2026, 10).[5] Suppose that's true. Still, most cognitive tasks intuitively face very steep diminishing returns or outright quality ceilings. The existence of these ceilings matters for the automation "barrier" that Gans and Goldfarb identify. The lower the quality ceiling, the lower the barrier to automation.

So, how low is the quality ceiling on cognitive tasks humans currently perform? Put otherwise: suppose a median- or extremely high-skill cognitive worker had *infinite* time to devote to some moderately or extremely demanding cognitive task. How much higher quality should we expect the

[5] Gans and Goldfarb have confirmed that their model can be interpreted in terms of the cognitive automation of cognitive tasks, though they don't think "full cognitive automation" is plausible (personal correspondence). Of course, that's what's at issue here! If the dynamics their model aims to characterize govern the current round of cognitive automation, then this thereby protects cognitive workers' wages (jobs) and, ipso facto, there's not "full" cognitive automation.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

output to be? Maybe twice for the median, and somewhat higher at the extreme, depending on the kind of task. Surely not more than one order of magnitude at any widely distributed skill level. If this is right, then an O-ring barrier to automation is a correspondingly easy barrier to clear, at least for the bulk of cognitive tasks.

2.6.6 Sources of Friction

The analysis so far assumes labor markets operate without major institutional interference and that standard economic dynamics largely control the outcome: firms substitute AI for human cognitive labor (almost) whenever doing so is cheaper (controlling for the opportunity cost of deploying the AI elsewhere). But labor markets don't operate in a vacuum. Various forms of friction might slow or redirect displacement even under continuation. Three kinds of friction are especially salient.

First, governments could restrict AI's role in cognitive work in various ways (e.g., via licensing requirements). This is the most immediate political lever available for stalling the displacement of human workers from cognitive work (Acemoglu and Restrepo 2020a). But regulatory intervention requires sustained political will against prevailing economic headwinds. Individual jurisdictions face competitive pressure compounded by automated cognitive work's lack of geographic constraints. Restrictive nations risk firms relocating cognitive tasks to lighter-touch jurisdictions, much as manufacturing was off-shored (Blinder 2006). In the absence of coordinated international action, regulation fragments and likely slows rather than outright prevents displacement.

Legal frameworks potentially create a second source of friction. Existing legal systems assign responsibility to corporations and to human beings. When AI systems make errors, liability can be uncertain: Does responsibility fall on the developer, the deployer, or the nominal human supervisor (Matthias 2004)? This uncertainty may make firms cautious about automation. But legal frameworks standardly adapt to economic pressure. In particular, new liability regimes emerge and insurance markets can be expected to develop in ways that permit automation to proceed (Trout 2025). Moreover, as Bayern (2016) has shown, it has been technically possible for *decades* to turn any instance of AI into a legally responsible entity without changing a single law.

Third, various groups of workers might organize collective resistance. But collective action faces significant defection problems. Individual firms or professionals who adopt AI in cognitive work are likely to gain

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

competitive advantage over those who resist. Some sectors with outsized cultural or political influence may hold out longer. For an example of each, SAG-AFTRA's recent strike provisions restricting AI use in production and longshoremen's decades-long resistance to port automation show that organized labor can slow adoption in specific sectors. But organized labor seems unlikely to be successful at halting the widespread economic transformation of work this time around, just as it was unsuccessful in halting previous technological transformations of work.

Indeed, all these dynamics have operated during earlier technological transitions. Craft guilds resisted mechanization in the nineteenth century. Regulatory frameworks struggled to accommodate factory automation. Liability questions slowed adoption of new industrial technologies: Railroad accidents in the nineteenth century drove the development of modern liability insurance. But in each case friction simply stretched the transition without stopping it. The same pattern seems likely here.

2.7 Summing Up the Case for Displacement

The case for displacement is conditional. If AI capabilities continue improving and costs continue falling at roughly current rates, then we can expect displacement to win the race with reinstatement decisively. Various equilibrium effects (may) explain why some human cognitive work will persist in a world where AI is increasingly capable. But these dynamics do not give us reason to think that the distinctive package of features defining cognitive work will survive an encounter with AI capable of reliably performing most (existing and future) cognitive tasks (Table 1).

A natural way to avoid this conclusion is simply to deny the relevant assumption: to predict that AI capabilities will plateau.

2.8 The Case for Ubiquitous Assistance

The case for widespread displacement rests on the assumption that AI capabilities keep improving. But what if capabilities plateau roughly where they are now?

Even under this conservative assumption, the status quo is not a plausible future. Current AI systems are already capable enough to transform cognitive work, and the forces driving adoption operate independently of further capability gains.

More or less widespread displacement is easy to picture. A world characterized by more or less ubiquitous cognitive assistance is somewhat hazier. There's plenty of talk about "working with AI," "co-intelligence,"

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Table 1 Each mechanism preserves some human work. None preserves the distinctive package of cognitive work.

Equilibrium effect	Mechanism	Why it fails under continuation
Demand expansion	Cheaper cognition expands demand	Demand saturates; not infinitely elastic
Task complementarity	Remaining human tasks become more valuable	Fewer workers needed, even if each is more valuable
Wealth effects	Productivity gains fund new employment	New work emerges, but not cognitive work specifically
Comparative advantage	Humans retain tasks where AI is relatively weaker	AI isn't scarce; human wages collapse toward AI costs
O-ring effects	Concentrated human attention raises quality	Quality ceilings are low; barrier easy to clear

"human–AI collaboration," and so on, but rarely much specificity about what's actually happening. What exactly does cognitive assistance *do*? How are we being "cognitively assisted"? Which parts of cognitive work get offloaded, and which remain human? And what likely follows for the package of features defining cognitive work?

2.8.1 What Is Cognitive Assistance?

In 1997, Garry Kasparov lost at chess to IBM's Deep Blue. Thereafter, Kasparov proposed formalizing a new format: "advanced chess," where human players could consult computers during matches. In advanced chess, human–computer teams ("centaurs" or "cyborgs") played better chess than either humans or computers alone. They divided the cognitive labor of chess: each side did what it was better at. Centaur chess is a paradigmatic case of technological assistance in the performance of a complex, multifactor, highly context-sensitive (though typically very low-stakes) cognitive task.[6]

[6] There are other examples: diagnostic decision-support systems exhibit a similar structure. But centaur chess is unusually clean: the division of labor is extremely legible, and the outcome is measurable. It's also instructive because of what happened next: as chess engines improved, the centaur advantage disappeared. Today, the best engines beat the

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

The centaur ("augmentation") model has therefore become the default frame for thinking about AI assistance in cognitive work (Acemoglu and Restrepo 2020a; Autor 2024; Lei and Kim 2024). The vision is appealing: AI handles the tedious parts of cognitive work while humans contribute judgment, creativity, and oversight. But "AI handles the tedious parts" is too vague to be analytically useful. Which parts, exactly, can we expect to be offloaded, and which are likely to remain human?

The answer turns on what makes cognitive work difficult in the first place: what we can think of as humans' *cognitive bottleneck*. As the case of chess nicely illustrates, humans face hard constraints on both the scale and the intensity of their effortful cognition. For example, we can hold only a few (dozen) items (e.g., states of the board) in working memory at once (Baddeley and Hitch 1974). Intensity, sustained attention (e.g., to a game) degrades over time (Monsell and Driver 2000), and switching between tasks (e.g., between move recall and move evaluation) incurs cognitive costs (Uncapher et al. 2016) that require recovery time to fully refresh.

More generally – that is, beyond just chess – effortful cognition requires exercising a cluster of related meta-cognitive capacities. This suite of capacities is what cognitive scientists call our "executive system," and it comprises "executive functions" that can only do *one thing at a time* and typically *not for that long* (though these constraints are liable to training) (Diamond 2013; Ahmed et al. 2019). For instance, capacities such as working memory, sustained attention, inhibitory control, and planning are among those that manage and coordinate conscious thought; these capacities all operate serially rather than in parallel (Miyake et al. 2000). Moreover, a well-functioning executive system is itself quite cognitively demanding to acquire and requires explicit, deliberate training (as compared to the implicit, informal training needed to develop cognitive capacities like language) (Center on the Developing Child 2011; Miyake et al. 2000).

Cognitive work is cognitively demanding precisely because it demands serial processing through this single-channel human cognitive bottleneck. For example, legal reasoning, programming, medical diagnosis, and strategic planning each require holding multiple considerations in

best human–computer teams. Assistance can be unstable: a waypoint to displacement rather than a permanent configuration. I return to this point in Section 4.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

mind, resisting multiple distractions, searching systematically through large information spaces, and monitoring each part of a long chain of reasoning for error. Even the most capable, best-trained humans can only do so much of this at once, and only for so long. (I return to cognitive difficulty and executive functions in Section 4.)

What, then, is artificial cognitive assistance? I propose the following functional definition: artificial cognitive assistance is assistance that aims to route around the human cognitive bottleneck in some way.

Given this definition, whether something counts as cognitive assistance depends on context. This variability is a feature of the definition: assistance is relational, indexed to the cognitive constraints of workers (e.g., particular workers on particular tasks).

Such assistance can take many different forms, depending on the details of the cognitive work in question, including the broader sector of work in which it is embedded. This is why, increasingly, companies are building *bespoke* cognitive assistants for different domains (Sharadin 2026b; Gabriel et al. 2024; Eloundou et al. 2023).

Return to centaur chess. The bottleneck framing explains why the human–computer division of labor takes the form it does. The goal is high-quality chess. The bottleneck is human cognition. The centaur configuration is how you route around it while keeping humans in the loop. Computers handle exhaustive search tasks that would overwhelm human limits. Humans contribute strategic judgment. Chess engines therefore count as cognitive assistants in our sense.

Now forget centaur chess. The same structure appears wherever AI assistance is being deployed in cognitive work: centaur *everything*.

In radiology, AI processes full scans and flags regions of concern–handling the systematic visual search that would otherwise load working memory and sustained attention. This frees radiologists to concentrate on evaluating the flagged candidates (Boeken et al. 2023). In legal research, AI searches exhaustively through case law while tracking relevance criteria, freeing lawyers to focus on evaluating which precedents actually matter for the case at hand. In software development, AI generates candidate code, handling the working-memory load of syntax and edge cases, freeing engineers to concentrate on architectural judgment and verification.

The subject matter varies, but the structure doesn't: AI takes on parts of tasks that would otherwise tax executive function, freeing humans to concentrate their limited cognitive resources on what remains.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

2.8.2 Why Will Cognitive Assistance Be Ubiquitous?

Cognitive assistance in cognitive work is assistance that aims to route around human cognitive bottlenecks in some way – the human–centaur X configuration, for whatever X is relevant. Why think this configuration will be increasingly commonplace? Why expect *ubiquitous* cognitive assistance?

Two forces, technological defaults and economic incentives, drive ubiquity, and together they create a ratchet.

Start with defaults. Defaults will structure people's choices about how (and when) to use this cognitively assistive technology in exactly the same way they systematically shape our choices in other areas.[7] Assistance will be the default. We know this because in domains where AI assistants are most mature, assistance is already part of the default interface for cognitive work. For instance, productivity suites have persistent "Help me X" for some more or less general cognitive X (e.g., write, generate ideas, draw, code) buttons. In some cases, users cannot adjust the default. This is in fact how the tools most workers use to complete cognitive tasks already now work, by default.

This emerging default status is not an accident; it is by deliberate design. Google's researchers are especially explicit about the future they're building; in a recent paper, they describe AI assistants as "the main interface of the future" (Gabriel et al. 2024, p. 15). The business models of dominant technology companies demand that AI assistance becomes the standard interface for cognitive work.

Moreover, defaults, as we know, can easily harden into norms (Bovens 2009; Thaler and Sunstein 2021). The social norm these companies seem to envision is one where unassisted cognitive work marks someone as an eccentric, like refusing to use a calculator, insisting on paper maps, or ignoring social media.

Economic incentives, meanwhile, accelerate these dynamics. Early adopters (driven by defaults) gain an edge; competitors adopt to keep pace; eventually nonuse marks a worker as incompetent. This is an arms race with a predictable outcome.

[7] Defaults are a powerful lever for shaping behavior. For an overview of recent research on this effect, see Jachimowicz et al. (2019). For a popular introduction to the use of defaults (and other systematic biases) in order to influence choice on matters of public policy – "nudges" – see Thaler and Sunstein (2021). For a philosophical overview, see Bovens (2009). For a qualified defense of nudging, see Levy (2019) and the more accessible overview in Levy (2018).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

In general, we can expect corporate actors to shape a path of least (in this case, cognitive) resistance and set it as the default; economic incentives will then ensure that people walk it. Some individuals will resist on principle. Some domains will remain stubbornly resistant to assistance. But the forces are strong enough to push most cognitive workers toward ubiquitous use. And as I'll now explain, ubiquitous use is sufficient to transform cognitive work.

2.8.3 How Will Ubiquitous Cognitive Assistance Affect Cognitive Work?

The most obvious impact of ubiquitous cognitive assistance on cognitive work is that it will disrupt our fifth feature: limited productivity growth. Centaur–Magnus can play many more simultaneous games at a higher level of quality than plain old human Magnus.

This is good news! Baumol's cost disease is a *disease*: nobody thinks it's a good thing that, for example, education and healthcare continue to grow relatively more expensive over time. Curing the disease means productivity can finally grow in sectors where it has historically stagnated.

Does this destabilize the other features defining cognitive work? Let's work through each.

High skill requirements will likely persist in an ubiquitously assisted world, since assisted work still requires expertise. Lawyers reviewing AI output must know law. Maybe more intuitively, I want my centaur-radiologist to be, *you know*, a well-trained, highly skilled radiologist! The specific skills will likely shift – more verification, less generation – but cognitive work is unlikely to become recognizably "low" skill. If it did, why wouldn't the already quite capable AI be doing it?

What about *credentials*? This feature likely faces significant disruption. Credentials currently strongly signal a capacity for, among other things, *generation*. For example, law school exams test brief-writing, medical boards test diagnosis, and programming interviews test code production. But assisted work is more likely to emphasize verification: evaluating AI output rather than producing it from scratch. For some period, credentials will therefore signal the wrong (at least one much less important) thing: skills correlated with what the assistant now handles rather than skills required for working effectively with assistance. Eventually, filtering and credentialing institutions such as universities will presumably adapt (cf. Arrow 1973).

In an ubiquitously assisted world, *automation resistance* is, of course, significantly weakened. But assistance might stabilize at a human-in-the-loop equilibrium. If, for example, verification genuinely requires human

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

judgment that AI cannot replicate, the centaur configuration could persist rather than sliding toward full displacement.

As I think of things, the crucial question is what happens with *wage premiums*. Productivity gains mean – potentially *much* – more output per worker. If demand for cognitive output doesn't expand proportionally (I'll say more about this below), then significantly fewer workers will be needed to meet demand. Significantly fewer workers needed to meet demand, same supply of trained workers, standard labor market dynamics: significant downward pressure on wages.

So the question isn't whether ubiquitous assistance would transform cognitive work: it would, by enabling significantly increased productivity growth in sectors that have suffered Baumol's cost disease for decades. The question is whether that transformation cascades into the package's other features – especially wage premiums – or whether it gets absorbed by countervailing forces. The next section considers reasons to doubt both the ubiquity of assistance and the severity of the cascade.

2.9 Resisting Ubiquitous Assistance

2.9.1 Denying Ubiquity

One might argue that capable assistive systems will exist but won't be ubiquitous, reducing both their positive impact on productivity and their negative impact on wages. I don't have much to say about this possibility.

Some cognitive work might escape assistance entirely. Perhaps there are some economically valuable tasks with no relevant human cognitive bottleneck to route around. I expect there are not many such tasks that currently exist (though see below on *new* tasks). Maybe tasks requiring judgments of taste are like this. But such tasks are typically embedded in workflows full of bottleneck-heavy cognitive tasks. Certainly, some institutions and sectors will resist adopting cognitive assistance for reasons beyond competitive pressure, and they will pay a premium (e.g., in growth or profits) for it. But if assistance becomes ubiquitous in *most* cognitive work while artisanal cognizers hold out, this will not preserve cognitive work as we know it.

2.9.2 Denying the Effect on Work

You might instead accept that assistance will be ubiquitous but resist the claim that it will have the two effects I've claimed it will on cognitive work. There are two salient possibilities.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

The first involves imagining worlds in which, although assistance is ubiquitous, it simply has no positive impact on productivity. For reasons I've already explained, this seems unlikely: These systems are *already* capable of performing a range of economically valuable cognitive tasks to a very high degree of reliability. Admittedly, this has not yet showed up in measures of per-worker productivity. But this is likely best explained by diffusion dynamics and well-known problems with measuring changes to per-worker productivity caused by changes in technology (Brynjolfsson 1993).

The second possibility is more interesting: productivity gains occur, but wages don't collapse. Several dynamics might produce this outcome.

First, consider the possibility of *new tasks* emerging. This is a standard economist's reply to pessimism about automation: technological change historically creates new tasks even as it automates old ones (Acemoglu and Restrepo 2019). But the relevant question here isn't whether productivity gains from automation will lead to the creation of new tasks, including tasks that require human labor. Suppose new tasks emerge broadly and quickly enough that unemployment stays below 10%. The relevant question here is whether these new tasks will be cognitive tasks that can be bundled into jobs that thereby preserve the distinctive package associated with cognitive work. Here I'm skeptical.

The mechanism that generates new tasks – roughly, productivity gains creating surplus, surplus funding new activities – doesn't preferentially create high-skill, credential-gated, wage-premium work, let alone work with those features that involves sitting in front of a keyboard. It creates whatever work the economy demands. If that's care work, personal services, entertainment, or some other new difficult-to-imagine kind of not-cognitive work, then new tasks will absorb the surplus of labor generated by ubiquitous assistance without preserving cognitive work as an economically relevant category.

I myself think we should be agnostic about the skill content, credential requirements, and wages of tasks that don't yet exist. By contrast, someone interested in resisting the effects of assistance on cognitive work who predicts that new tasks will emerge to absorb productivity gains needs to explain why *cognitive tasks in particular* will be the ones that so emerge. Crucially, the burden of proof is then on this person to do something that economic history teaches us it's extraordinarily difficult to do: they must predict the specific skill content of new tasks (Autor et al. 2003).

Now consider the possibility of *demand expansion* in a world of ubiquitous assistance. As we saw earlier, in worlds of displacement, expanded

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

demand for cognitive output can simply be met with more AI. But in a world of ubiquitous assistance, expanded demand will by hypothesis require more human-AI "centaurs" – which, of course, means more humans. For example, suppose cheaper legal analysis leads to *vastly* more demand for legal analysis that requires human verification. Then, productivity gains might translate into more assisted lawyers rather than fewer well-paid ones. Whether this dynamic effectively preserves cognitive work depends on two things. First, assistance must remain stable rather than sliding toward displacement. Second, demand elasticity must remain high enough that expanded output absorbs the productivity gains. Both of these possibilities are very uncertain, but neither is crazy. I lack the space to consider them in detail here.

Finally, consider the effect on cognitive work of ubiquitous assistance in *noncompetitive sectors*. Large portions of cognitive work occur in sectors insulated from competitive pressure to minimize labor costs. For instance, education, healthcare, government, and nonprofits employ substantial fractions of cognitive workers. But wages in these sectors are typically set by, for example, politics, collective bargaining, and institutional norms rather than marginal productivity calculations. Nevertheless, it seems unlikely that these sectors can preserve cognitive work's wage premiums indefinitely if the broader cognitive labor market collapses.

2.10 Against Status Quo

A status quo future for cognitive work is not a serious possibility. The only way to preserve the existing package of features that currently defines cognitive work – high skill requirements, credential-based gatekeeping, wage premiums, historical resistance to automation, and limited productivity growth – is to believe that AI systems that *already exist* won't systematically reshape cognitive work. That looks increasingly like wishful thinking in the face of forces already in motion (Table 2).

The most likely outcome, of course, is some mixture of outcomes: displacement in some domains, assistance in others, with variation by task, sector, and skill level. High-end cognitive work requiring judgment in novel situations might remain human while routine analysis is automated, or creative tasks might stay human while verification and compliance are displaced. The precise distribution is impossible to predict. But the conceptual tools developed here can help us navigate questions that emerge in mixed worlds.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Table 2 Both scenarios break the productivity constraint. Displacement eliminates the remaining features; assistance shifts, distorts, or erodes them.

Feature	Status quo	Displacement	Assistance
Skill requirements	High; years to develop	Eliminated; AI performs tasks directly	Shifted; generation → verification
Credentials	Gate entry; signal expertise	Eliminated; nothing left to certify	Distorted; signal wrong capacities
Wage premiums	High; reflect scarcity of skills	Eliminated; supply vastly expands	Eroded; productivity outpaces demand
Automation resistance	Strong; nonroutine work is hard for machines	Eliminated; AI targets nonroutine work directly	Eroded; human-in-loop may hold
Productivity growth	Stagnant; bottlenecked by human attention	Rapid growth; bottleneck removed	Growth; bottleneck routed around

In my view, the precise mixture of displacement and assistance matters much less than a point that the standard normative framing gets backward. Widespread displacement of humans from cognitive work is typically supposed to be a serious threat to human flourishing. Ubiquitous assistance, meanwhile, is supposed to be the solution. The next two sections argue the reverse: displacement poses no special danger to human flourishing, while assistance carries a distinctive developmental risk we have largely ignored.

3 Widespread Displacement

3.1 Introduction

Section 2 argued that the widespread displacement of humans from cognitive work is one of two serious possibilities for the future of cognitive work. Displacement certainly *sounds* alarming. This section argues that it may not be as bad as it sounds; it may be quite good.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

To make the argument as difficult as possible, I'll focus on the extreme case: near-total displacement, a world where almost all cognitive work is performed by artificial systems rather than humans. If even this extreme scenario poses no (special) threat to human flourishing, then less extreme scenarios involving only partial displacement (which are more likely) probably don't either. I return to this inference in Section 3.5.

The argument has two parts. First, I argue for a *parity claim*: cognitive and not-cognitive work are on a par with respect to the (noneconomic) goods of work. In other words, there's nothing special about cognitive work from the standpoint of human well-being (again, controlling for actual wage patterns). So, if displaced cognitive workers shift to not-cognitive work, they haven't lost access to whatever goods work provides. Second, I argue for a *productivity claim*: displacement of humans from cognitive work in particular (as opposed to work in general) enables massive, leveraged gains in domains where such gains tend to disproportionately enhance human welfare.

I then address three objections to this optimistic view of displacement: that there won't be enough not-cognitive work to absorb displaced cognitive workers, that the productivity gains from displacement will be unjustly distributed, and that displacement would entail an objectionable loss of human control over our collective future. I'll argue none of these is a special worry for displacement per se.

3.2 The Parity Claim

The parity claim says that cognitive and not-cognitive work are on a par with respect to the (noneconomic) goods of work. If the parity claim is right, then displaced cognitive workers who shift to not-cognitive work haven't lost access to whatever welfare goods work provides. Instead, they've merely changed the way in which they access these goods.

I'll offer three arguments for the parity claim. The first is empirical: the goods typically associated with work in fact appear in both cognitive and not-cognitive work. The second is conceptual: there's nothing about cognitive work being "cognitive" that could make cognitive work specially valuable. The third is theoretical: no major theory of well-being identifies cognitive work as distinctively good for people.

3.2.1 The Empirical Argument

What makes work good for people, assuming it sometimes is? Here, we can safely ignore the effect of money, since there's obviously nothing about

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

cognitive work per se that guarantees better pay than not-cognitive work. The usual (nonpecuniary) candidates for the goods of work are things such as engagement, achievement, social connection, autonomy, and purpose (cf. Crawford 2009; Danaher 2017; Gheaus and Herzog 2016; Cholbi 2023). If cognitive work had special access to these goods, that could explain why the displacement of humans from cognitive work is a distinctive threat to human flourishing. But it doesn't. Each of these goods appears in not-cognitive work.

Consider engagement first. Csikszentmihalyi's research on flow states found deep engagement as readily in carpentry as in coding (Csikszentmihalyi and Csikzentmihaly 1990; Csikszentmihalyi et al. 2014). The absorption that makes work satisfying doesn't track the cognitive content of tasks. Instead, it tracks the match between challenge and skill (Csikszentmihalyi 1975).

The goods associated with achievement follow the same pattern. A master craftsman achieves as genuinely as a master analyst (Sennett 2008). For instance, an electrician might troubleshoot a baffling fault by exercising judgment. This counts as achievement on even the most demanding accounts (Bradford 2015; Hurka 2020).

Social connection forms around shared challenges and mutual reliance, neither of which requires specifically cognitive tasks. Workplace bonds on a construction site are at least as deep as those in consultancies.

Autonomy is often cited as the most important good (other than money!) associated with work (cf. Gheaus and Herzog 2016). But cognitive work has no monopoly on autonomy. Instead, worker autonomy varies more by job structure than by the cognitive content of job-related tasks. A tenured academic has autonomy; a junior associate billing hours at a partner's direction doesn't. Journeyman plumbers have it; call center workers don't. Autonomy at work intuitively tracks hierarchy and independence, not the cognitive content of job-related tasks.

What about purpose? I doubt this is a construct for which we have a good empirical measure: purpose is woefully under-specified as an empirical construct (cf. Rosso et al. 2010). (I return to this point in Section 5.) But intuitively, purpose, if anything, favors not-cognitive work. It would be strange to claim that plumbing or nursing involves less purpose than consulting or copywriting.

The goods that make work good (when it is) appear across both cognitive and not-cognitive work. (Notably, so do the *bads*: alienation and tedium afflict cognitive and not-cognitive workers alike.)

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

3.2.2 The Conceptual Argument

Perhaps the empirical argument misses something. Maybe there's a conceptual route to the conclusion that humanity's displacement from cognitive work would be a special loss.

If cognitive work is uniquely valuable for human flourishing, then something beyond its economically distinctive features must make it so. What could that be? A natural idea: cognitive work involves thinking, and thinking is arguably distinctively human. So perhaps a loss of cognitive work is a loss of an activity that's especially human. Let's grant that thinking is in some way especially human. Still, this doesn't explain why a loss of cognitive work is distinctively bad.

This is because not-cognitive work involves a lot of thinking, too. A carpenter deciding how to approach a tricky joint is engaged in planning; a mechanic reasoning backward from symptoms to causes is engaged in problem-solving. The difference between cognitive and not-cognitive work tends to lie in the *object* of thought, not in the presence or absence of thought per se.

At this point, a defender of cognitive work's special status might try a different move. Perhaps what makes not-cognitive jobs good, when they are good, is precisely their cognitive elements. The carpenter's job is good for them, or promotes their living a flourishing life, because of the planning and problem-solving it involves – that is, the cognitive parts of the job. The "not-cognitive" label that we apply to these jobs is in this respect misleading; the especially valuable core of these jobs is cognitive after all.

This has a whiff of circularity. It amounts to saying: a specific kind of cognitive activity is what makes work valuable; so, whenever we have the intuition that allegedly not-cognitive work is valuable, that must be because of some of that work's (latent) cognitive components. But this begs the question. Why think planning is what makes carpentry valuable (pleasurable, contribute to flourishing, and so on) rather than its physicality? The move also carries an unpleasant odor of elitism – as if the philosophers and analysts get to claim the good parts of manual labor as secretly their own, while the sawing and hammering are mere dross.

3.2.3 From Existence to Parity

The arguments so far establish that the goods typically associated with work *exist* in not-cognitive work. But, one might object, existence is strictly weaker than parity. There are at least two relevant possibilities.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

First, perhaps these goods are systematically *less accessible* in not-cognitive work. Maybe flow states are rarer on construction sites than in design studios, or the conditions for autonomous judgment are harder to achieve when work is physical. This is an empirical question, and as we've seen, the evidence doesn't support the asymmetry (Dormann and Zapf 2001; Humphrey et al. 2007).

Second, and more interestingly, one might argue that different kinds of work have *internal goods* specific to the practice (MacIntyre 2007). On this view, what makes legal reasoning valuable for a lawyer isn't generic "achievement" but something intrinsic to the practice of law – a form of excellence that can't transfer to plumbing, however excellent the plumbing. Displacement from one practice to another might therefore involve genuine loss even if both practices offer engagement, autonomy, and the rest.

This objection has real force, but it cuts against special treatment of cognitive work. If internal goods are practice specific, then displacement from *any* practice to *any* other involves losing access to the relevant practice-dependent goods. A displaced carpenter loses the internal goods of carpentry just as a displaced lawyer loses the internal goods of law. This is a reason to care about displacement from specific practices generally, not a reason to think displacement from cognitive work is distinctively threatening. Moreover, the internal goods of a practice are goods *of the practice*, not straightforwardly goods *for the practitioner*. The excellence internal to law matters to law; whether it matters to the lawyer's welfare is a further question that the generic goods are better positioned to answer.

3.2.4 The Theoretical Argument

We can confirm the parity claim by examining major theories of well-being directly. Hedonists care about pleasure, which, as we've seen, tracks challenge–skill match rather than cognitive content (Gregory 2015; Dormann and Zapf 2001). Preference theorists defer to what people actually want, and preferences about work are diverse – where preferences for cognitive work exist, they typically turn out to be instrumental, tracking wages and working conditions rather than cognitive activity as such (Heathwood 2015; Deci and Ryan 2000). Objective-list theorists point to goods such as achievement, knowledge, and autonomy (Fletcher 2015) – goods the empirical argument has already shown appear in both categories. No major theory of well-being identifies cognitive work as distinctively good for people.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

3.2.5 Summing Up the Case for Parity

Each route leads to the same place. The goods associated with work appear empirically in both cognitive and not-cognitive work (Table 3). There's nothing conceptually special about "cognitive" that could plausibly ground an evaluative distinction between cognitive and not-cognitive work. And no major theory of well-being identifies cognitive work as intrinsically better for people. The parity claim holds: cognitive and not-cognitive work are on par with respect to the (noneconomic) goods of work. For this reason, the widespread displacement of human beings from cognitive work is not, *as such*, a loss.

3.3 The Productivity Claim

The parity claim is defensive: it shows why displacement isn't a special loss for human flourishing. The productivity claim is positive: displacement could be a special boon to human flourishing.

3.3.1 Research and Discovery

Section 1 identified limited productivity growth as a distinctive feature of cognitive work. As we've seen, Baumol's cost disease makes cognitive-labor-intensive goods steadily more expensive because productivity can't keep pace with wage growth (Baumol 1967, 2012).

Table 3 The goods that make work good for people are explained by mechanisms indifferent to cognitive content.

Good of work	Mechanism	Cognitive example	Not-cognitive example
Engagement	Challenge–skill match	Coding	Carpentry
Achievement	Competent causation under difficulty	Master analyst	Master craftsman
Social connection	Shared challenges and mutual reliance	Consultancy teams	Construction crews
Autonomy	Job structure	Tenured academic	Journeyman plumber
Purpose	Tangible contribution to others' needs	Legal Aid	Plumbing, nursing

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Cost disease isn't the only constraint on productivity growth afflicting sectors that rely on cognitive work. Research productivity *per worker* has been declining for decades: The number of researchers required to achieve a given rate of productivity growth has increased dramatically over time (Bloom et al. 2020; Gordon 2016; Kortum 1997). Ideas, as Bloom et al. put it, are getting harder to find.

These two constraints compound each other. Cost disease means that the cognitive labor required for research is expensive and trending more so. Declining research productivity means we need even more of that expensive labor to achieve the same results.

The widespread displacement of humans from cognitive work attacks both constraints, though in different ways.

Replacing human cognitive labor with AI attacks cost disease directly by massively expanding the effective supply of cognitive labor.

Displacement may also attack declining per-worker research productivity. The fact that ideas are getting harder to find arguably reflects the increasing difficulty of navigating complex search spaces. As fields mature, the unexplored territory that might contain breakthroughs becomes harder to identify and reach. Human researchers, with finite careers and cognitive limitations, can only explore so much (Bloom et al. 2020).

AI systems may be differentially well suited to this kind of exploration. Consider drug discovery. Traditional pharmaceutical research involves scientists generating hypotheses based on existing knowledge and intuition, then testing candidates through slow (and expensive) experimental pipelines (DiMasi et al. 2016). Systems such as diffusion models for molecular design do something different: They explore chemical space in ways humans couldn't, identifying candidate molecules that human researchers would never have considered (Watson et al. 2023).

These systems don't themselves displace researchers – they don't use themselves! But if displacement proceeds as Section 2 suggests and capabilities continue to improve, the systems doing the exploring will increasingly be the ones deciding *what* to explore *by themselves using other AI tools*.

Google's "AI co-scientist" (Gottweis et al. 2025) offers a glimpse of what this future might look like. Consider the example from Google's paper introducing the system: A team of researchers recently spent several years piecing together the solution to a biological puzzle involving bacterial parasites and arrived at their breakthrough only through what they themselves described as serendipity. Before their results were published, Google gave the same puzzle to co-scientist, which they had designed

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

to mirror the kind of cognitive labor scientists do when developing hypotheses about novel biological phenomena. The system's top-ranked hypothesis was *identical* to the human team's breakthrough (Penadés et al. 2025). Importantly, this is a first-generation system: To my knowledge, it's the first attempt by *anyone* to build a system designed to displace human beings from these kinds of complex cognitive tasks.

Breaking Baumol's cost disease gives us more cognitive labor at the frontier; there will be more "minds" working on problems, without the training pipeline that currently limits supply and without the associated costs of employing human researchers. This is CEO of Anthropic Dario Amodei's "country of geniuses in a datacenter" (Amodei 2024). But if the frontier is genuinely hard, if good ideas require searching vast spaces that human cognition isn't well suited to navigate, then more human-style cognition will also hit diminishing returns. We may need not just more cognition but different kinds of cognition. AI that explores differently, not just faster, could steepen the curve of discovery itself (Cockburn et al. 2018).

The welfare implications of this are significant. Research generates compounding returns: breakthroughs enable further breakthroughs (Jones 2005; Romer 1990). And these benefits persist through time: a new class of antibiotics saves lives for decades. Accelerating research is about as high leverage an intervention for improving long-run human welfare as it's possible to imagine.

3.3.2 Access to Justice

There's a separate reason to welcome displacement: in many societies, completing certain cognitive tasks is a prerequisite for fair institutional treatment, and automating those tasks would democratize access to justice.

Legal services are the obvious case. If legal tasks are expensive, justice is correspondingly expensive to access. The poor and otherwise vulnerable therefore predictably get less access to justice (Rhode 2004; Sandefur 2019). Displacement from legal labor therefore matters because fair institutional treatment shouldn't depend on ability to pay.

The kinds of cognitive tasks involved in tax preparation and benefits administration illustrate the same point. The American tax code – like many (but not all) tax codes in developed economies – is baroque (Hoppe et al. 2023). Completing the cognitive tasks required to conform to it in the most efficient way demands either significant expertise and the time to deploy it or the money to hire an expert's time. The same is true of

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

applications for various kinds of government programs: They tax applicants' time and discourage eligible people from claiming what they're entitled to (Lowrey 2021; Herd and Moynihan 2019). And in these cases, the largest burdens of compliance fall hardest on those with the fewest resources to shoulder them. Intuitively, wealthy individuals have relatively complex taxes but are eligible for fewer benefit programs with high administrative burdens; by contrast, poorer individuals have relatively simple taxes but are eligible for relatively more high-burden benefit programs.

We could address these problems by eliminating the relevant administrative "sludge" (Sunstein 2021, 2022). But that requires a very specific kind of political action. We could also subsidize human help, for example, by expanding legal aid and funding "benefits navigators" or providing free tax preparation services. But human cognitive labor is expensive and scarce.

Assistance could perhaps ease these constraints. But even dramatically more productive pro bono lawyers can only serve so many clients, and even dramatically cheaper lawyers are still too expensive for many who need help. Systems capable of displacing human cognitive workers, meanwhile, change the economics qualitatively. When AI performs these tasks *without human cognitive labor*, the marginal cost of helping one more person shoulder some unfair administrative burden is compute and electricity.

3.4 Objections

3.4.1 Not Enough Work

The most commonly cited worry about displacement of humans from cognitive labor is that there won't be anywhere for displaced cognitive workers to go. Previous technological transitions had destinations: agricultural workers moved into manufacturing, and manufacturing workers moved into services and cognitive work. But cognitive work was supposed by many to be the final destination – the refuge for workers displaced from everything else (Mokyr 2011). If it's automated too, where do people go?

There are two versions of this worry.

The first is that there simply isn't enough *not*-cognitive work to absorb displaced cognitive workers. This is very uncertain, though I'm inclined to be optimistic. Not-cognitive work as we've defined it here includes the skilled trades, personal services, entertainment, physical labor in unstructured environments, large portions of healthcare, and much else. Whether this sector can absorb a large influx of former cognitive workers depends on elasticity of demand, the pace of transition, and policy choices

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

about work-sharing and labor markets (Acemoglu and Restrepo 2020a; Autor 2024).

There are hard questions about how to make the transition to not-cognitive work easier for individual cognitive workers. But I doubt we should be concerned that there won't be "enough" work to go around. The "lump of labor" fallacy is supposed to be a *fallacy* precisely because new work typically emerges as old work is automated (Autor 2015). Whether this pattern continues under the AI-driven displacement of cognitive work is uncertain. But if it does, then the concern is about managing the transition smoothly, not about displacement being bad as such. Faster transitions mean more acute disruption, but the disruption is the same kind previous generations faced. And we have better tools now than, for example, nineteenth-century agricultural workers had; in particular, we have safety nets and the resources to fund them (Esping-Andersen 1990). The question is whether we deploy these tools adequately, which is a political question, not a reason to think displacement from cognitive work into not-cognitive work is uniquely threatening.

A second version of this worry is that future AI systems will eventually displace *not*-cognitive work too, foreclosing lateral moves by workers entirely. This version of the worry is significantly more speculative. In Section 2 I noted that embodied tasks in unstructured environments remain harder to automate than those that involve pure cognition. But set that aside. Suppose embodied AI systems (e.g., humanoid robots) eventually displace humans from not-cognitive work as well. Then the concern is that humans have nowhere left to sell their labor. I'll say two things about this more speculative possibility.

First, it seems obvious to me that the mere fact that no entity is economically incentivized to purchase any human's labor is not itself a special moral reason for concern. Instead, this tends to be morally concerning when, for instance, this fact attends various problematic distributions of resources. I address such worries about distributive justice below.

Second, I'm tempted to go further. The argument for the parity claim aimed to show that not-cognitive work offers the same menu of goods arguably required for human flourishing as cognitive work offers. But it also implied something stronger, namely that work (in the economic sense) isn't strictly necessary for human flourishing. After all, there's no reason to think that *being paid* is a necessary condition for enjoying any of the (noneconomic) goods that we noted are typically associated with (both cognitive and not-cognitive) work (things such as achievement, social connection, and so on).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

In any case, notice that in this second version of the worry, the concern is no longer about displacement from cognitive work specifically, but about displacement from work as such. That concern turns on a different question – whether performing remunerated tasks is necessary for human flourishing. I am extremely skeptical, but addressing this issue in any detail would go beyond the scope of the present discussion.

3.4.2 Distributive Justice

Someone might admit that displacement *could* be good but only if the benefits associated with displacing human cognitive workers from cognitive work are shared. Unfortunately (the thought goes), they won't be. Capital will predictably capture all (or most) of these benefits; meanwhile, displaced human workers (and their families) will be immiserated. I'll make two initial comments designed to allay the concern, and then explain why I think this is not best understood as a concern about displacement per se.

First, the natural distribution of displacement's benefits deserves careful attention. One of these benefits is potentially very meaningful productivity gains in sectors currently suffering from Baumol's cost disease. Take one such sector: medical research. Productivity gains in medical research yield better medicine. And the benefits of better medicine don't flow primarily to capital; they flow to patients. For example, a new cancer treatment primarily benefits patients who have cancer, not shareholders, even if the treatment is sold on the open market for a profit. This remains true even if the opportunity to access the treatment is deeply inegalitarian: Welfare gains from medical research aren't captured by capital in the way profits are. Similar remarks apply to other sectors afflicted by low productivity. The sludge-clearing benefits noted above are even more clearly egalitarian: The whole point is that administrative burdens fall hardest on those with the fewest resources, so the immediate effect of removing them is to help the worst-off most.

The distributive worry makes most sense for productivity gains that show up as corporate profits. It makes much less sense for gains that show up as, for example, cures, or as the removal of barriers to fair treatment.

Still, the concern isn't baseless. Some of the gains from automating human cognitive labor will almost inevitably take the form of profits, and the distribution of those profits is arguably (Nozick 1974) a legitimate concern. But, and this is my second comment, this represents a familiar problem of distributive justice. In fact, it's almost exactly the same problem of distributive justice that arose with industrialization, with globalization,

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

and with *every previous wave of automation*. We have myriad frameworks for thinking about what a fair distribution of the surplus generated by cooperation requires (Rawls 1971; Arneson 1989; Parfit 2018; Frankfurt 2018) and many known mechanisms for achieving it (James et al. 2022; Van Parijs and Vanderborght 2017). The question is whether we use them.

The structure of this objection is revealing. It concedes that the widespread displacement of human beings from cognitive work *plus* a fair distribution of the benefits thereof yields a good outcome. It then predicts that for some reason we won't achieve such a fair distribution. But that's a (pessimistic) political prediction, not a claim about displacement itself. If the prediction is right, the correct response is to work harder on distributive politics, not to prefer a world where displacement doesn't happen and the relevant gains don't exist to be distributed. In general, the correct response to a bad distribution of some good is a better distribution of that good.

So, at least as I understand things, there's nothing about displacement causing unjust distributions that makes those distributions a morally distinctive problem *for displacement itself*. Consider an analogy. Suppose a company invents a drug that makes humans effectively immortal. Left to natural market forces, this would presumably have very seriously inegalitarian effects on the already very unequal global distribution of wealth. But it would be very misleading to put this point by saying things like: there's a serious moral problem with making humanity immortal. Equally, in the case of widespread displacement.

3.4.3 Autonomy and Distribution

Suppose some distribution is adequate in material terms and meets whatever further evaluative standard (if any) we think an economic distribution must meet. Still, something troubling could remain. Under wage labor, people's material welfare depends on their ability to sell something others need. This arrangement has many problems. But it arguably gives people a kind of independence. If displacement extends beyond cognitive work to labor generally – or if the not-cognitive sector can't absorb displaced cognitive workers – then people's welfare will depend on the continued willingness of whoever controls productive capacity to share it. Perhaps there's something problematic about this kind of dependence. I am skeptical, for three reasons.

First, the autonomy of wage labor is easy to romanticize. The freedom to sell your labor provides little independence if you'll starve without a

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

buyer; in that case, unrestrained buyers will dictate terms. And buyers – that is, employers – are in fact largely unrestrained (cf. Anderson 2017). This is why most workers have a right of exit from their employment in name only. Dependence on labor markets may therefore be no better than dependence on distributive institutions.

Second, the worry here is really about institutional design, not (even *total*) displacement as such. Concentrated ownership of AI *plus weak institutions for governance* could indeed leave people more vulnerable than they are now. But that's an argument for better institutional design and better governance, not an argument against displacement itself. There's no shortage of proposals for improving governance along whatever dimension of freedom we think matters; these face political obstacles, but they're not conceptually mysterious.

Third, dependence *as such* is arguably not worth caring about. For example, republican political philosophers such as Philip Pettit argue that what makes dependence objectionable is *arbitrariness* (Pettit 1997). The relevant question on this view isn't whether we depend on others (or on institutions) but whether that dependence is structured by accountable institutions or left to the discretion of the powerful (Pettit 2012).

3.4.4 Loss of Control

Kulveit et al. (2025) argue that even incremental AI displacement could permanently erode humanity's influence over systems we depend on. Let's grant that this result would be bad. Is this a reason to think total displacement of humans from cognitive work would be bad? No. Let me explain.

Societal systems currently align with human interests partly through explicit channels (voting, consumer choice) but more fundamentally through their implicit dependence on human participation. For instance, states need citizens to generate taxable income, and markets need workers to produce and consumers to buy. This dependence creates alignment between human preferences and social systems not by explicit design but because systems that fail to support human flourishing lose the human participation they require. Widespread displacement threatens to sever this link. If AI generates most economic value, the historical connection between taxation and representation weakens. If AI handles most labor, markets optimize for AI activity rather than human preferences. The systems keep functioning – but they no longer need us, and so they no longer need to serve us: we are, in Kulveit et al.'s evocative phrase, "like cattle in an industrial farm – fed and housed by systems they neither comprehend nor influence" (Kulveit et al. 2025, 5).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Being treated the way cattle are treated in industrial food production is, again, an admittedly very worrying possibility! (Notably, consensus on this point is some reason to think we are failing in our moral duties to cattle raised for slaughter.) But notice what is required to arrive at this outcome: economic displacement broad enough that human participation becomes *completely dispensable* for economic production in roughly the same way the cattle's preferences are treated as completely dispensable in terms of our decisions about how to treat them. (I set aside the cultural dimensions Kulveit et al. discuss, since economic displacement does the argumentative work.) This is an extremely speculative possibility. In particular, it is significantly more speculative possibility than the possibility I'm focused on here in this Element, which is (more or less) widespread displacement of human beings from cognitive work in particular. Cognitive work is roughly half of employment in developed economies, and the other half promises to be substantially harder to automate. Displaced cognitive workers who move into not-cognitive work still pay taxes and still spend their wages on goods they want. Hence I am relatively sanguine about the likelihood of Kulveit et. al's extremely grim outcomes in the case of humanity's widespread (even total) displacement from *only* cognitive work in particular. Again, things may look different in outcomes where *all* economically valuable work is automated.

If a residual worry involving a loss of control remains, I suspect it concerns accountability rather than the distribution of human-versus-artificial cognition as such. If AI systems controlled by a narrow elite making consequential judgments without democratic oversight would be troubling, then *human elites doing the same should be equally troubling*. My own view is that both these outcomes are less troubling than many think, for broadly meritocratic-cum-perfectionist reasons (cf. Brennan 2016; Bai 2019). But set that to one side. The controlling question, if we are worried about this somewhat different issue, is whether crucial decision-making (human *or* artificial) is subject to whatever level of accountability we think is in fact appropriate – and if it isn't, how we can ensure that, going forward, it will be. That's a question about collective governance on which reasonable people disagree – not a reason to think displacement is per se problematic.

3.5 Conclusion

The widespread displacement of humans from cognitive labor is perhaps not the normative disaster it first appears. Whatever makes work good for human flourishing is equally available in not-cognitive work. And displacement promises substantial welfare gains.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

The section has focused on near-total displacement of humans from cognitive work to make the argument as difficult as possible. But one might worry the relevant inference doesn't run cleanly in reverse – in other words, that *partial* displacement could be worse than *total*. A total transition from driving on the left to driving on the right is fine; a partial transition is carnage! Could scenarios involving displacement be like that?

In the case of displacement, I think there's little reason for this concern. For partial displacement to be uniquely problematic, there would have to be costs that attend partial but not total transition. These could be of roughly three kinds: coordination costs, mismatch costs, or transition costs.

Coordination costs arise when everyone must move together for anyone to move in a welfare-positive way (as in the driving case). But labor markets don't require simultaneous action. A cognitive worker and a former-cognitive-worker-now-plumber aren't in the same market; nothing forces their choices to align. Mismatch costs arise when being in different systems creates friction or danger for those who remain. But a world with half as much cognitive work isn't more dangerous to the remaining cognitive workers the way a road with half the cars on the wrong side is dangerous to everyone. Transition costs are, as we've mentioned, very real. But these scale with the number of people transitioning. Partial displacement involves a smaller transition than total displacement. So, if total displacement is acceptable after its transition, partial displacement is acceptable after its smaller one.

If displacement dominates, human flourishing need not suffer. Indeed, it may substantially improve. The next section argues that ubiquitous assistance, the scenario that sounds gentler, may actually be worse.

4 Ubiquitous Cognitive Assistance

4.1 Introduction

In Section 1, I introduced the idea of cognitive work as defined by a package of five features. Section 2 argued that even extremely conservative assumptions about the trajectory of AI systems entail that the status quo for cognitive work is untenable. We face a future characterized by widespread displacement or ubiquitous assistance or, more likely, some mix of the two.

In Section 3, I argued that even total human displacement from cognitive work may not be as threatening as it sounds. Whatever makes

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

cognitive work good for human flourishing is available in not-cognitive work; displacement enables potentially massive gains in research and in people's ability to access fair treatment under the law; and standard reasons for pessimism about displacement identify familiar (primarily political) challenges rather than novel moral problems.

Ubiquitous assistance, meanwhile, can intuitively seem like a welcome alternative to widespread displacement. Workers mostly keep their jobs, things go on largely as before, but workers' lives get cognitively easier. It's natural to think this is an unalloyed improvement to the status quo. Help is a good thing! Ubiquitous cognitive assistance from AI systems promises help at *unprecedented scale*. How could that be bad?

This section argues that a world filled with ubiquitous cognitive assistance may have several bad features. The argument turns on a simple observation: cognitive assistance works by reducing cognitive difficulty – the degree of cognitive effort required to accomplish a cognitive task. That's just what "help" means in this context. If an AI assistant isn't making cognitive tasks easier, then it's failing at its intended purpose. But cognitive difficulty can be good in three distinctive ways: It can be a source of pleasure, it is arguably a constituent of achievement, and, most importantly, the experience of cognitive difficulty is necessary for the normal development of foundational cognitive capacities.

4.2 What Is Cognitive Difficulty?

Cognitive difficulty has both a subjective and an objective component. The subjective component is familiar: the "feeling of mental effort" that we have when (say) doing a crossword (Bermúdez 2025; Székely and Michael 2021). The feeling of effort is typically aversive (Inzlicht et al. 2018; David et al. 2024). This helps explain why we avoid cognitive effort whenever possible, for instance, by preferring to rely on "fast and frugal" heuristics (cf. Gigerenzer and Goldstein 1996) rather than slow and costly reasoning.

But cognitive difficulty isn't merely a feeling. It also has an objective dimension: the degree of load on the capacities involved in effortful, deliberate thought. Not all cognitive processing loads these capacities. The vast background cognition that hums along below conscious awareness – what psychologists call "System 1" cognition (Wason and Evans 1974; Evans 2008) or Kahneman's "thinking fast" (Kahneman 2011) – leaves them largely untaxed. That's why we "cognitive misers" prefer System 1 when we can get away with it – it's experienced as largely effortless (Stanovich

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

2020). Cognitive difficulty involves "System 2" thinking: slow, effortful, and deliberate cognition.

Effortful cognition imposes a range of managerial demands on our cognitive systems. Cognitive scientists call the system handling these demands our *executive system* (Baddeley and Hitch 1974; Diamond 2013). The particular executive functions that compose this system include: inhibitory control (suppressing distractions, resisting impulses) (Munakata et al. 2011); working memory (holding and manipulating information in mind) (Baddeley 2020, 2007); cognitive flexibility (shifting perspectives, adjusting to new information) (Braem and Egner 2018; Ionescu 2012); planning and organization (sequencing actions toward goals) (Miyake et al. 2000); and error-monitoring (checking reasoning and action for error as it unfolds) (Yeung and Summerfield 2012). A well-functioning executive system enables the effortful cognition that complex thinking involves.[8]

This framework lets us understand cognitive difficulty more precisely as *load on the executive system*. A task is cognitively difficult to the extent it taxes these specific cognitive capacities: requiring sustained attention (inhibitory control), demanding that you hold multiple considerations in mind simultaneously (working memory), and so on.

This conception tracks our intuitive judgments about the cognitive difficulty of tasks. Merging onto a busy highway is cognitively difficult because it requires rapid attention-switching, inhibition of the impulse to hesitate, and continuous updating of your mental model of surrounding traffic. The same framework explains why some jobs are regarded as cognitively demanding: these jobs involve *bundles* of tasks that load executive functions heavily.

This conception is empirically robust: load on executive functions is the standard opera-tionalization of task difficulty across cognitive psychology, neuroscience, and educational research. When researchers assess

[8] The System 1 / System 2 distinction has a long history; versions appear in James (1890) and were formalized by Wason and Evans (1974), and it has proven productive across cognitive and social psychology (Evans 2008). Researchers disagree about whether the two "systems" are genuinely distinct neural substrates or merely different processing modes of a unified cognitive system. Nothing here turns on this question. What matters is our ability to distinguish the two kinds of processing by their hallmarks: S2 cognition is controlled, systematic, analytic, explicit, rule-based, and reflective (Evans 2008, 257). Second, the executive functions I focus on follow Miyake et al. (2000); see also Diamond (2013), and Doebel (2020). Some researchers distinguish "lower-order" executive functions (e.g., working memory, inhibitory control) from "higher-order" ones (e.g., planning) that recruit them (Diamond 2013). Since nothing I say depends on this hierarchical organization, I set it aside.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

whether one task is harder than another, they measure demands on working memory, inhibitory control, and the rest (Diamond 2013; Miyake et al. 2000).

With this conception in hand, we can ask: what could be *good* about cognitive difficulty? This question admits different answers depending on *whose* cognitive difficulty we're considering. Adults, whose executive capacities are largely formed and relatively stable (Zelazo et al. 2004), face one set of considerations. Children, whose capacities are still developing, face another.

4.3 Why Difficulty Matters Prudentially for Adults

Cognitive difficulty is arguably prudentially good for adults in two ways: it can be enjoyable, and it may be partly constitutive of achievement.

4.3.1 Pleasurable Difficulty

Cognitive difficulty is typically aversive (Székely and Michael 2021). But under the right conditions – particularly when the activity demanding the cognition is freely chosen (Wahlers et al. 2025) – it can also be enjoyable. Rawls articulated a general version of this idea as the "Aristotelian Principle": "other things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realized, or the greater its complexity" (Rawls 1971, 373). The principle applies to capacities of all kinds, including our *cognitive* capacities. We often enjoy the effortful engagement of our (innate or trained) cognitive capacities.

This enjoyment appears in cognitive work when conditions permit. The key empirically robust condition is a balance between challenge and skill (Csikszentmihalyi and Csikzentmihaly 1990). When challenge is too low, boredom results; when too high, anxiety. The sweet spot, where difficulty stretches but doesn't exceed capacity, produces the absorbed enjoyment that Aristotle noticed, Rawls codified, and Csikszentmihalyi popularized.

How could ubiquitous cognitive assistance threaten this value? Easy: assistance could systematically shift much cognitive work below the relevant difficulty threshold for absorbed enjoyment. As a result, cognitive work could become systematically more *boring*.

4.3.2 Difficult Achievement

Cognitive difficulty might also be good for a more theoretical reason. Gwen Bradford has argued that "genuine" achievement requires two

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

things: competent causation and difficulty (Bradford 2015). On Bradford's view, you achieve some outcome just in case you skillfully control the causal process that produces it *and* it was difficult for you to do so.

A lawyer who researches, analyzes, and drafts a complex brief has achieved something in Bradford's sense. An assisted lawyer who prompts an AI system, reviews its output, and approves the filing has certainly done *something*. But has she *achieved* anything? Maybe not. If difficulty is constitutive of achievement, then reducing the difficulty of some task reduces the degree to which completing that task is an achievement. At the limit, if cognitive assistance eliminates the cognitive difficulty associated with a task entirely, then completing that task may not count as an achievement at all.

Why care whether assisted cognitive workers "genuinely" or "Capital A" (cf. Bradford 2015) achieve (as much)? Achievement appears on nearly every plausible objective list of welfare goods (Fletcher 2015; Parfit 1984). If less difficulty means fewer achievements, then *ceteris paribus* assistance makes assisted lives go worse by depriving people of a widely recognized component of human flourishing.

There are obvious ways to resist this worry. One is to reject Bradford's view – to deny that difficulty is constitutive of achievement. I'm sympathetic to this move on theoretical grounds: It lets us say the ubiquitously assisted achieve just as much as the unassisted, only that they do so *effortlessly*.[9] But this doesn't dissolve the worry about value; it merely relocates it. Even if effortless achievement counts as genuine achievement, it may not contribute to welfare in the same way. For theorists who think achievement is a welfare good, what matters isn't the achievement itself but the effort behind it. Simon Keller is explicit on this point, saying that a "plausible rough-and-ready" claim relating achievement to welfare is that "the greater the effort required for an individual to achieve her goal, the more her welfare is enhanced by its achievement" (Keller 2004, 34).

On this view, then, reducing the effort required for achievement *ipso facto* reduces its contribution to welfare (whether or not it reduces the extent to which it is an achievement). Since ubiquitous assistance will *by hypothesis* reduce the effort required for a wide array of (more or less cognitive) achievements, it will thereby reduce the contribution these achievements make to individuals' well-being.

[9] For discussion, see Hirji (2019) and Sharadin (2026a).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

4.4 Why Difficulty Matters Morally for Children

Consider a thought experiment. Imagine parents who provide their child with ubiquitous *physical* assistance. Whenever the child might have to expend physical effort in order to accomplish a task, an adult physically intervenes to reduce the physical difficulty associated with completing that task. The parents are loving and attentive; they simply want to spare their child the demands of physical effort. The result is predictable: such a child would fail to develop normal motor control and coordination. Basic physical skills that require practice would be severely enfeebled (Adolph and Franchak 2017; Franchak and Adolph 2024).[10]

Now transpose this thought experiment to cognition. Imagine a child raised with ubiquitous *cognitive* assistance. Whenever the child might have to expend cognitive effort to accomplish a task, an ubiquitously available artificial system intervenes to reduce the difficulty. The result is just as predictable: such a child would fail to develop normal cognitive control and coordination, which is to say they would fail to develop robust executive function. Basic cognitive skills that require practice would be severely enfeebled.

Developmental psychology confirms this commonsense intuition. Contrary to popular "myth," the executive functions "do not automatically develop with maturity over time" (Center on the Developing Child 2011).[11] Instead, they develop through regular *use*; in particular, they develop through repeated encounters with cognitively demanding tasks that load these capacities (Best and Miller 2010). Children learn to hold multiple things in mind by doing things that require holding multiple things in mind. Children learn to inhibit impulses by being in situations where impulse must be inhibited. Cognitive difficulty is a necessary condition for normal cognitive development (Burns et al. 2026). This is because opportunities to exercise (or develop) executive function depend on experiencing cognitive difficulty.

[10] The converse finding (that increased movement opportunities accelerate motor development) is documented in (Lobo and Galloway 2012). For evidence of severe motor delays resulting from extreme movement restriction in institutionalized children, see Beckett et al. (2006); note that these latter cases involve broader deprivation beyond physical assistance alone.

[11] This claim is repeated throughout the literature on executive function and cognitive development (cf. Best et al. 2009; Zelazo et al. 2004; Best and Miller 2010). Philosophers from a very diverse variety of traditions make the same point about the development of the "virtues" of practical and theoretical reason (cf. Aristotle 2014; Schapiro 1999; Giesinger 2012; Tan 2017).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Children don't seek out cognitive difficulty. Instead, like adults, they're cognitive misers who prefer the path of least resistance. But difficulty is embedded in the environments children frequently navigate: in particular, in schoolwork requiring sustained, deliberate cognition. A ten-year-old doesn't *want* to plan an essay; she wants to be done with the assignment. But the only path to being done runs through actually planning it. This is what aligns opportunity to practice cognition with the incentive to do so. Ubiquitous cognitive assistance in educational contexts disrupts both sides of this equation. It creates a new path of least resistance – one that routes around difficulty rather than through it. And it thereby undermines children's existing incentives to engage in developmentally beneficial cognitive difficulty.

This isn't armchair speculation. After I drafted this section, Brookings' Center for Universal Education released a report that puts the point starkly:

> [Students'] brains are developing, undergoing crucial processes of neural pruning and strengthening that depend on repeated cognitive effort and struggle. They lack the metacognitive skills, critical thinking abilities, and neurobiological maturity of adults. School exists precisely to build these capacities through sustained engagement with challenging material. The productive power of AI tools, which amplifies adult expertise, can undermine this developmental goal when placed in the hands of learners. Cognitive development requires the effortful processing, mistake-making, and problem-solving that AI too easily circumvents and that many students are willing to bypass. Used liberally, AI is not a cognitive partner; it is a cognitive surrogate. It does not accelerate children's cognitive development–it diminishes it. (Burns et al. 2026, 57)

Worse, these dynamics (and the resulting developmental harms) are unlikely to be experienced equally. Instead, like most harms to children, they will be regressively distributed.

One might object that this concern is primarily a worry about certain uses of technology in educational contexts rather than a concern about this technology's impact on children's cognitive development *as such*. But children face cognitive challenges outside school, and we already have a model for what it looks like when a ubiquitous assistant routes around them: the so-called helicopter parent. Helicopter parenting has been consistently linked to lower well-being in children and young adults, and – crucially for our purposes – to significantly worse outcomes *specifically for executive function* (Moilanen and Lynn Manuel 2019; Schiffrin et al. 2014).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

The mechanism is precisely what we'd expect: The parent who plans, organizes, remembers, and intervenes is using *their* executive function so that the child doesn't have to exercise her own. Predictably, this leads to children (and young adults) who struggle to form plans, organize their lives, and generally get on with the business of life (McCoy et al. 2025). Of course, even the most devoted helicopter parent faces constraints. The enfeebling impact of her cognitive assistance is therefore limited by the fundamental scarcity of human attention. Ubiquitous artificial cognitive assistants face no such constraints.

Ubiquitous cognitive assistance therefore threatens children's cognitive development on two related fronts. First, it threatens to reduce the rate at which children confront meaningful cognitive demands in their environments both in and outside educational contexts. If capable AI assistants are increasingly capable of handling cognitively demanding tasks, and these systems are ubiquitous in environments in which children would otherwise face demands on their own cognition to complete those tasks, then children may simply encounter fewer situations that in fact load their executive functions. Second, it threatens to undermine the incentives children have to engage in whatever cognitively demanding tasks remain: When assistance is ubiquitously available, why bother to struggle cognitively?

The core concern for children, then, is that this simultaneous reduction in opportunity and incentive leads to (more or less widespread) arrested cognitive development, specifically when it comes to their executive systems. (I note that analogous concerns may also apply to children's emotional development [Burns et al. 2026]; here, I leave these troubling issues to one side.) Intuitively, this is a very serious moral problem. I'll say more about the philosophical background for being especially morally concerned about threats to children's cognitive development below.

Before that, let me anticipate an objection. Why are we suddenly talking about children in a book about cognitive work? Children don't typically do cognitive work in the economic sense. The answer is that cognitively assistive systems don't have natural boundaries.

These tools are being deployed as the basic infrastructure for accomplishing cognitive tasks (Section 2), and infrastructure designed for adult purposes routinely reshapes childhood even when children aren't its intended users. Cars are a nice example. Cars made suburbs possible, and car-dependent suburbs made children dependent on adults for

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

mobility. One arguably bad result of this was the chauffeured childhood of scheduled playdates and organized activities. Smartphones are another case in point. Smartphones were designed for adult communication and productivity, but their ubiquity has radically transformed childhood sociality and, arguably, children's basic attentional capacities (Haidt 2024).

Ubiquitous AI assistance can be expected to follow this pattern, but with a troubling difference. Cars weren't built to be driven by children and, thankfully, they still usually aren't; smartphones weren't initially designed for use by eight-year-olds, though, unfortunately, they now increasingly are. In both cases, the reshaping of childhood was largely incidental. With the emergence of capable AI assistance, by contrast, things are much more deliberate. Systems initially built to solve problems of adult productivity are already being intentionally integrated into ubiquitous educational tools as defaults (Miller and Mervosh 2025). Schools are rushing to adopt AI to prepare students for an AI-saturated economy (Burns et al. 2026); public policy is pointed in this direction (cf. Trump 2025); and corporations interested in locking in the next generation of users are pushing on this open door.

But perhaps this overstates the concern. Even if ubiquitous cognitive assistance impedes the development of executive functions, why think this is an especially serious *moral* problem? After all, these capacities are valuable in part because they are useful for navigating a cognitively demanding world. In an assisted world, maybe they matter (much) less. If AI handles planning, organization, and sustained attention, perhaps weaker executive functions are simply an adaptation to new circumstances, much like how we've outsourced memorization (and much else) to our phones without obvious catastrophe.[12] Moreover, perhaps the overall welfare gains from ubiquitous cognitive assistance will outweigh whatever developmental costs it imposes on children.

This response misunderstands what our executive systems are fundamentally *for*. These capacities aren't useful tools for navigating cognitively demanding environments that we might discard when our environments become less cognitively demanding. They are capacities constitutive of autonomous agency itself: They enable us to formulate, revise, and pursue our "conceptions of the good" (cf. Rawls 1971). And so a person with impaired executive function doesn't just struggle with cognitive tasks; they struggle, much more fundamentally, to direct their own life in accordance

[12] But see Haidt (2024). I say more about the "extended mind" below.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

with a reflective understanding of their own values. The capacities at stake are in a very real sense the psychological infrastructure of autonomous self-governance.

Even granting this, one might reply: So what? Perhaps some people would prefer to live assisted lives with a diminished capacity for autonomy. Who are we to judge? One answer would be to simply insist that these capacities are in fact intrinsically good, even if one doesn't care about them (cf. Giesinger 2012; Kant 1785). A more ecumenical answer is that whatever adults might permissibly choose for themselves, the moral status of decisions that risk threatening the normal cognitive development of children are fundamentally different.

Philosophers who otherwise disagree about much else converge on the following idea: Adults owe children special duties precisely because children are developmentally immature – beings who are in the process of becoming autonomous agents.[13] And despite their theoretical differences, these views all share a commitment to what we can call the Open Future Principle:

> **Open Future Principle:** It is *pro tanto* morally wrong to impede a young person's normal development of the cognitive capacities required for autonomous agency without substantial justification.[14]

Notice that this principle is very weak in three ways. First, it's *pro tanto* rather than absolute; so, countervailing considerations might justify developmental costs in particular cases. Second, it requires only "substantial" justification, not decisive proof. Third, it protects only those capacities foundational to autonomous agency; so, not every cognitive skill that might be useful, or that we might value for some other reason, is covered by it.

Yet even in this extremely weak form, the Open Future Principle has teeth. In the present case, it explains why a world characterized by ubiquitous cognitive assistance poses a very significant moral risk: such a world, as we've just discussed, threatens to impede children's development of the cognitive capacities required for autonomous agency.

[13] For a selection of views, see Gheaus et al. (2019), Gheaus (2021), Schapiro (1999), Feinberg (1998), Brighouse and Swift (2014), Franklin-Hall (2013), Roth and Surprenant (2012), Jaworska and Tannenbaum (2018), and Brennan and Noggle (1997).

[14] Of course, I borrow the term "open future" from Feinberg's (1998) influential defense of children's "right" to an open future. But the principle as stated is much weaker than Feinberg's view; as I say, I know of no view concerning adults' moral relationship to children that doesn't accept it.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

To my knowledge, no one building or deploying these systems has offered a substantial justification for imposing this risk on a generation of children. Equally, I am aware of no evidence that providing ubiquitous cognitive assistance to children *improves* developmental outcomes when it comes to children's executive function. Quite the opposite. Moreover, plausible justifications for building and deploying cognitively assistive technology in a way that affects children are likely to concern adult productivity (conceiving of children largely as future workers who must be prepared for an AI-saturated economy) and adult convenience (appealing to the efficiency of educational models that rely on AI tutors). These may be real goods! But we've been given no empirical reason to think a childhood characterized by ubiquitous cognitive assistance actually produces them. And more importantly, even if it did, these goods wouldn't obviously outweigh the potential developmental harms.

4.5 Objections

I'll now consider reasons one might think these concerns with an ubiquitously cognitively assisted future are less serious than I've suggested.

4.5.1 Extended Mind

Maybe the framing of cognitive assistance I've offered is confused. Cognitive processes aren't confined to the skull; instead, our cognition extends into our tools and environments. On this view, AI assistance is simply the latest extension of our minds. There's therefore no "offloading" (Risko and Gilbert 2016) to worry about, no substitution of external for internal, just cognition distributed differently across a hybrid human–AI system (Clark 2010). Steve Jobs famously described computers as "bicycles for the mind" (Agrawal et al. 2025). Maybe worrying about cognitive ease brought on by ubiquitous cognitive assistance is like worrying that your legs are doing less work because you're riding a bicycle. I have three responses to this line of thought.

First, what AI assistants do looks more like *substitution* than *extension*. The paradigm cases of extended cognition involve tools that augment capacities the agent is actively deploying: in Chalmers and Clark's (1998) famous example, a notebook augments the memory of an Alzheimer patient, Otto. In this example, Otto is actively using his memory (albeit in diminished form) to store and retrieve items from the notebook.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

AI assistance, meanwhile, often simply replaces the agent's cognitive engagement entirely. This is especially true when it's being used not by adult professionals to enhance their productivity in high-skill cognitive work but by children to avoid the aversive feelings associated with cognitive difficulty (Burns et al. 2026).

Crucially, this is not an accident: substitution *rather than extension* is the stated aim of the largest companies developing cognitively assistive technology. For example, researchers at Google explicitly disavow the extended mind framing, saying that "we are suggesting that AI assistants are not merely external cognitive subsystems that constitute part of the user's extended mind" (Gabriel et al. 2024); later, they double down on this characterization, saying that the AI assistants Google aims to develop and deploy to as many individuals as possible are *not* properly analogized to "the notebook in the Clark and Chalmers [. . .] example of an Alzheimer's patient who uses a notebook as a functional substitute for biological memory" (Gabriel et al. 2024, 15). I suggest we take the designers of these tools at their word.

Second, even if we accept the idea that this technology "extends" rather than substitutes for our own cognition, this doesn't resolve the concerns I've identified. The nature of our *un*extended cognitive capacities can still matter a great deal! An extended cognitive system with a well-functioning human component differs in significant ways from the same system with a degraded human component. This is much like how a cyclist with strong legs differs from one with atrophied legs, even if both get around equally well *in some environments*. Equally, if the human substrate of executive function simply works worse in the future, then even if a *perfectly* ubiquitously cognitively assisted person may not experience this difference, they will still be significantly cognitively enfeebled *when unassisted*. We may care intrinsically about this fact, as many moral philosophers no doubt do (cf. Kant 1785). Or we may care instrumentally about it, since on occasions where assistance is less than perfectly ubiquitous, this may cause significant harm.

Third, and crucially, when it comes to children, it is, as we've seen, morally illicit for adults to do things that close down opportunities for children to freely make the choice to "extend" themselves. To freely make this choice, children must be in possession of the fundamental cognitive capacities to make it. But the present trajectory is toward structuring the developmental environment in ways that make it *less* rather than *more* likely that children will in fact develop these capacities, and so eventually be in a position to freely choose for themselves whether they want to

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

be "extended." In my view, this is a bit like Clark and Chalmers *giving* poor Otto early-onset Alzheimer's, buying him a notebook, and calling it a wash.[15]

4.5.2 Difficulty Is Often Bad

One might object that I've romanticized cognitive difficulty. Much cognitive work is tedious and alienating. Why mourn its passing?

It's correct that not all difficulty is valuable. Stressful cognitive overload doesn't build executive function. The kind of difficulty that develops our cognitive capacities is *appropriately challenging* engagement. As I noted earlier, this can be understood broadly as exposure to cognitive tasks that stretch our capacities but don't overwhelm them. Tedium and burnout fail this condition from opposite directions.

But ubiquitous cognitive assistance doesn't selectively remove "bad" cognitive difficulty while preserving "good" cognitive difficulty. Capable AI assistants that can perform cognitive tasks more or less chirpily handle *whatever* cognitive tasks they can handle. This increasingly includes the challenging, judgment-requiring parts of cognitive tasks, which are precisely the parts that constitute appropriately demanding engagement. For adults, this may mean what remains is the residue AI can't yet touch: for instance, a final sign-off that carries legal liability. But this isn't a part of cognitive work that produces flow or constitutes (welfare-enhancing, "genuine") achievement. Of course, as we've noted, adults can presumably find cognitive engagement elsewhere; so maybe this isn't so prudentially bad, on net.

For children, I've argued the moral risk is more serious: the existing balance of opportunity and incentive for productive cognitive struggle is poised to shift dramatically. The existing equilibrium is one where children are regularly presented with many situations that demand cognitive struggle *and* where they are incentivized to struggle cognitively in those situations in part because it's difficult to avoid a satisfactory result without doing so. These are the cases of (developmentally) "good" difficulty.

But what does an ubiquitously assisted future look like for children? It's one where *by definition* fewer situations in children's cognitive

[15] Extended mind theorists can of course acknowledge all this (Clark 2003). Their thesis largely concerns the boundaries of cognition, not whether some components of the unextended system matter more than others for autonomy and development.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

environments meet these two conditions – there are fewer demands and weaker incentives. Crucially, this is true *even if no one optimizes for this environment*.

Consider, again, the analogy to our physical environments. Cars, elevators, and suburban infrastructure weren't designed to make children sedentary. But they helped make childhood in developed economies substantially more sedentary nonetheless. A child in 1926 walked to school, climbed stairs, and played in the street; a child in 2026 is driven everywhere, takes escalators, and mostly plays indoors.[16] The physical environment became easier to navigate without anyone intending to reduce the opportunity or incentive for children to engage in physical activity.

Instead, the reduction in opportunity and incentive for exerting physical effort was a side effect of infrastructure built for adult convenience. The same applies here. Cognitive assistants are built to make cognitive tasks easier for adults in the same way escalators are built to make stairs easier for adults. But a system designed to help lawyers draft briefs can obviously help a twelve-year-old write an essay. To imagine otherwise requires imagining either that children somehow won't have access to the defining technology of their era (we'll ban children from escalators) or that they'll systematically choose the cognitively more demanding path when an easier one is available and unstigmatized (they'll freely choose stairs). As I'll now explain, neither idea is plausible.

4.5.3 Restricting Access to Children

Maybe you think the developmental concerns are serious but tractable through straightforward policy: restrict children's access to enfeebling cognitive assistance until their executive functions are sufficiently developed. If ubiquitous assistance threatens childhood development, then we should treat it like we treat alcohol or tobacco (ideally also: smartphones and social media) and impose age restrictions, require parental controls, and enforce robust educational guidelines.

I have four comments on this idea.

First, implementing policy in this area requires settling on a principled age range. The developmental science points somewhere uncomfortable. The prefrontal cortex continues developing into the mid-twenties (Fuster

[16] Admittedly, children in 1926 may also have worked at a factory. The analog in the present case would be cognitively demanding child labor–which, to be clear, I am not currently advocating! The point isn't that 1926 was better for children overall; It's that physical demands were embedded in daily life in ways that have since been engineered out.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

2002). Capacities such as impulse control, long-term planning, and cognitive flexibility are still being refined when people are graduating college, entering the workforce, and starting families (Smith and Jonides 1999). A restriction that actually tracked executive function maturity would need to extend well beyond what we legally or culturally consider childhood.

Of course, this is an old challenge that we've managed to solve for other developmentally harmful technology. So suppose we agree on an age. Still (my second comment), *enforcement* of an age restriction in this case is extremely challenging. One natural thought is that schools are the lever: ban cognitive assistance in educational settings and the problem is largely solved. But this overestimates how much of children's cognitive development happens in school. Children spend only roughly 15% of their waking hours in formal education (Walberg 1984); the rest is homework, play, socializing, navigating the world, and–increasingly–interacting with digital infrastructure that contains ubiquitous cognitive assistance by default. If executive functions develop through repeated encounters with cognitive demands across contexts, then protecting the classroom while leaving the rest of childhood cognitively frictionless accomplishes relatively little.

The broader point is that AI assistance doesn't arrive as a discrete product that can be age-gated at point of sale like tobacco or corralled onto dedicated platforms like social media. It's embedded in digital infrastructure. Meaningfully restricting access would require coordinated action across institutions on a scale unheard of in the history of technology governance. For the record, I am largely in favor of attempting this. But it's a big ask!

Third, even if technically enforceable, the political economy includes a bevy of perverse incentives. Suppose that we don't get global coordination on restrictions for children's use of these systems. Then the question is how various actors who could in principle impose meaningful restrictions are incentivized to introduce such restrictions.

Consider first the incentives facing schools and parents. Schools adopt AI tools because parents demand them, because governments encourage "future-ready" curricula (cf. Trump 2025), and because the tools appear to improve measurable outcomes such as test scores and assignment completion rates (Burns et al. 2026). Whether these gains reflect genuine learning won't be clear for years. Education research is notoriously slow, and controlled studies that might reveal long-term developmental effects lag far behind adoption curves. In the meantime, a school that restricts AI while competitors embrace it risks losing students to districts that seem more forward-looking. Parents, meanwhile, face their own competitive

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

pressures: withholding assistance from your child when peers have access looks like handicapping them for college admissions and career opportunities. And these incentives for parents to demand more AI in schools feed back into school policy, either directly or indirectly (e.g., through politics).

What about governments? Most municipalities lack the capacity to set enforceable restrictions on any kind of technological diffusion; worryingly, some nations lack this capacity too. Among those capable of restricting access to new technology, competitive pressures push in the opposite direction. A country that limits AI in education while rivals embrace it risks falling behind in the global contest for economic and technological leadership. Restriction looks like unilateral technological disarmament. Of course, it's only *childhood* disarmament – sounds good to me! But perhaps unfortunately, given the globally competitive landscape, it seems unlikely that most governments will agree to coordinate on this kind of restriction. This pessimistic conclusion is compatible with some governments, following Australia's leading moral position on a symmetrical set of issues regarding social media (Nogrady 2024; UNICEF Australia n.d.), acting unilaterally to help their own country's children. It is also compatible with multi-lateral agreements aimed at reducing catastrophic risks from advanced AI systems.

Fourth, this idea – ban AI assistance for children – isn't really an objection to the claims I've argued for here in this section: it instead concedes the core concern. If keeping cognitive assistance away from children is a good idea, then the developmental risks identified in this section are serious enough to warrant significant formal or informal restrictions on children's access to a technology currently being built and deployed at enormous scale and in basically unrestricted fashion. To repeat: I am myself in favor of such restrictions! But as I currently understand the policy landscape, this would be a very significant shift from current policy and practice. Current policy and practice point toward *increasing* rather than *decreasing* the exposure of children to cognitive assistance.

4.5.4 People Will Choose Difficult Things

In an assisted world, people can seek cognitive difficulty voluntarily, either in cognitive work itself or outside such work. Sports, hobbies, puzzles, and games provide cognitive challenge without economic pressure. Those who value the goods provided by cognitive difficulty can pursue these goods deliberately; those who don't can enjoy the ease (cf. Danaher 2017, 2019). What's wrong with letting people choose? This is an optimistic picture.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

For adults, there are two problems with this optimistic picture. First, ubiquitous cognitive assistance at work doesn't magically free up time for demanding pursuits elsewhere. A (full-time) job that's been cognitively hollowed out is still a (full-time) job. Cognitive demands may shrink, but we've been given no reason to think the hours will also shrink, and economic history suggests they won't (Autor 2015). (Unless, of course, total work hours shrink to *zero*. But that is a different concern.) Cognitive workers won't be liberated to spend their newfound mental energy on challenging hobbies; they'll simply be left with less cognitively demanding work that fills the same hours in an assisted world as it did in an unassisted one. Second, we can expect that in a cognitively assisted world choosing cognitive difficulty *within* cognitive work will be very strongly penalized.

For children, the optimistic picture flatly misrepresents the moral situation. While it could possibly be reasonable to blame an adult for choosing cognitive ease over cognitive effort in some contexts, this is unreasonable when it comes to children. Instead, the amount of cognitive effort it makes sense to expect children to exert is largely a function of the cognitive demands they face in their environments. These environments, in turn, are shaped by the choices of adults – parents and, to a much lesser extent, schools.

4.5.5 Executive Function Won't Matter in Assisted Worlds

If AI handles the cognitively demanding parts of life, perhaps skill at executive function is simply less valuable than we thought. There are two versions of this objection, depending on what kind of value is at issue.

One idea is that executive function is less *instrumentally* valuable in a less cognitively demanding world. People arguably don't need navigation skills in a world with GPS; maybe people don't need planning and organization in a world with ubiquitous cognitive assistance.

To this I say: That depends on how ubiquitous the assistance is and which other features of the world are held fixed. If we imagine a world of extreme abundance – such as Pixar's WALL-E, maybe – then perhaps the cognitive demands on individuals are truly minimal, and executive function is instrumentally worthless. If we imagine (what is more likely) a world where although work is transformed, there's still "moderate scarcity" in material terms (cf. Rawls 2001), then I expect there will still sometimes be significant instrumental returns to having robust executive function. Here's a simple argument: even the *most capable*, *most ubiquitous* cognitive assistant can't stop a teenager from doing something rash. A well-functioning executive system can.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

A second, different idea is that executive function is less *intrinsically* valuable in a less cognitively demanding world. But this seems confused. Among those who think the capacities involved in executive function are intrinsically good to possess, this is usually said to be because those capacities (in whole or in part) constitute a capacity for autonomy, which is itself supposed to be intrinsically good (cf. Schapiro 1999; Roth and Surprenant 2012). But a capacity for executive function wouldn't cease to be a constituent part of a capacity for autonomy simply because a machine could in principle do all of our hard thinking for us. So, on the view that executive function is intrinsically good for people, it's difficult to see how it could be less so in an ubiquitously assisted world.

4.5.5.1 You're Saying This Time It's Different

Writing offloaded memory; calculators offloaded arithmetic; GPS offloaded navigation. Each time, critics worried about harm to our cognitive capacities, and each time humanity cognitively adapted without obvious cognitive catastrophe. Everyone always says that *this time it's different*. But why should a world characterized by ubiquitous cognitive assistance of the sort you've described be significantly different from previous worlds characterized by other kinds of cognitive assistance in the form of things such as writing, the calculator, and GPS?

I do think that the present case is importantly different from the cases of writing, the calculator, and GPS. And because those three examples are regularly mooted as alleged counterexamples to the claim that AI seriously threatens our cognition, it's worth spending a bit of time with this objection. I will say two things in response, one glib and one much more serious.

The glib thing to say is that I think, once we look at the data, the historic critics of technology seem to have been correct to be concerned in all three of these cases, for roughly the reasons they cited at the time they were concerned. Plato's Socrates famously worried in the *Phaedrus* about how the spread of literacy would affect cognitive skills relevant to the preliterate tradition (including memory) (Plato 1997). He was right: People really are measurably worse at remembering things in literate as compared to oral societies, and new technology continues to exacerbate that effect (Grinschgl et al. 2021; Sparrow et al. 2011; Uncapher et al. 2016). Teachers in the 1980s worried about students' ability to do mental arithmetic as calculators spread to classrooms (*Christian Science Monitor 1986*). And they were right: People really are worse at mental arithmetic (Ellington

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

2003), and numeracy rates are steadily falling (National Assessment of Educational Progress 2022). Finally, as someone who has long been worried about people's spatial cognition as a result of ubiquitous GPS-enabled smartphones (personal correspondence), I am pleased to report that *I* am right. People who rely on their smartphones for navigation really are worse at spatial reasoning of many different kinds (Dahmani and Bohbot 2020; Berki 2022). The track record of people worried about technology's effect *on our cognition specifically* is in fact surprisingly strong.[17]

But the objectors' real point is presumably that humanity *adapted* to past cognitive offloading. This deserves a serious reply: the present case poses special challenges because what's being offloaded is categorically different.

The adoption of earlier technologies involved offloading specific first-order cognitive skills. But it left the meta-cognitive skills comprising our executive system intact, sometimes even strengthened. For instance, the spread of literacy enabled us to offload the longterm storage and retrieval of information. But *writing well* in fact *loads* executive function heavily: It requires planning, sustained attention, error-monitoring (Ruffini et al. 2024). This is why teaching children to write well is a large part of how we train children's executive systems to function well (Burns et al. 2026). Similarly, the diffusion of calculators enabled individuals to offload arithmetic computation, a narrow skill that's largely automatic once mastered. But using calculators to solve problems requires (among other things) knowing which operations to perform and why, and strongly *loads* executive function (Ellington 2003). Again, this is why learning arithmetic is valuable even if one never does math outside the classroom: it helps train executive function. Equally, the advent of modern GPS (and especially ubiquitous GPS in the form of the smartphone) enabled individuals to offload their spatial reasoning, but deciding where to go, and adjusting plans when circumstances change, all executive functions, remained with the user.

In each case, the management layer of cognition–the executive system–stayed in the loop; in some cases (as with math and writing), its use *increased* as we tasked other elements of our cognition to our tools. This is no coincidence. When you offload a specific first-order cognitive skill to a tool, you create a new coordination problem: you must manage the

[17] Though it's not without its outliers! My favorite example is that people in nineteenth-century England were worried about something known as "railway madness," which could (allegedly) be caused simply by riding (at the time, quite noisy) trains (Hayes 2017).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Table 4 Previous cognitive technologies offloaded first-order skills while preserving or increasing executive function demand. AI assistance targets the executive layer itself across domains.

Technology	Skill offloaded	Executive function demand
Writing	Memory	↑ Increased
Calculators	Arithmetic	↑ Increased
GPS	Navigation	→ Maintained
AI assistance	Planning	↓ Decreased
AI assistance	Writing	↓ Decreased
AI assistance	Research	↓ Decreased
AI assistance	Summarization	↓ Decreased
AI assistance	Error-checking	↓ Decreased

interface, decide when to deploy the tool, interpret its outputs, and integrate them into your larger goals. That coordination work *is* the job of our executive systems. The tool handles execution; you handle management. So offloading execution often increases managerial load.

This dynamic is why our adaption to past cognitive technology has been so successful; in some ways, it can seem inevitable that we would adapt to these tools, given enough time.

The argument for inevitable adaptation runs as follows. Our executive systems comprise the suite of capacities for flexibly adapting to changing environments. That is, as we've discussed, what the executive system is *for*, in a neurofunctional sense: We cognitively adapt *by engaging our executive systems*. For example, we adapted to writing by first *learning* to read and write – both activities that strongly load executive function – and subsequently, you know, *actually* writing and reading – both *also* high-load activities. Since we weren't offloading our capacity for cognitive adaptation itself, adaptation was strongly favored (Table 4).

But it would be extremely foolish to take this precedent as evidence that, this time, it's not different. This is instead extremely strong evidence that, this time, it's different in a very worrying way.

In the present case we're facing a threat to our capacity *for cognitive adaptation itself*. That meta-cognitive capacity is obviously more fundamental to successful adaptation than any first-order cognitive skill.

Adaption to earlier rounds of cognitive offloading to technology was, if not inevitable, then at least extremely likely. This is because these technologies didn't target our flexible capacity for cognitive adaptation – our

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

executive system – itself. Quite the opposite: they tended to load that capacity quite heavily. But, this time, things are different in at least two ways. First, the technology – capable AI assistance – is *directly* targeting for offloading (planning, reasoning, and so on) things we do only by exercising our executive systems. And, second, there's no corresponding extra demand being automatically induced elsewhere on our executive systems by the offloading (as happened in the case of literacy).

To this we can add two points. First, the track record suggests that we typically "adapt" to the advent of cognitive assistance with respect to some cognitive capacity (e.g., longterm memory, mental arithmetic, spatial reasoning) by large swaths of people simply *abandoning* the cultivation of that cognitive capacity altogether. Abandoning the capacity for cognitive adaption would be, in my view, a bad idea.

Second, an interesting analog is the case of technology's effect on our attention (arguably an executive function; Watzl 2011; Monsell and Driver 2000). But the dynamic with, for example, smartphones and attention has not been actually been a matter of our offloading this cognitive skill; it's been more like technology's *fracking* it (Hayes 2025). And if anything, this recent case study should make us *even more worried*, since we arguably haven't adapted to the recent technologically induced disruptions to our attention.

4.6 Conclusion

This section has argued that ubiquitous cognitive assistance poses serious risks to human flourishing. The severity and type of risk differ for adults and children.

For adults, the risk is prudential and seems relatively moderate. Cognitive difficulty can be a source of absorbed enjoyment and may be partly constitutive of genuine achievement. Ubiquitous assistance threatens both: work that's been cognitively hollowed out may fall below the threshold where engagement becomes absorbing, and achievements produced with substantial AI help may not count as achievements in the robust sense that welfare theorists care about. One obvious response to this is: Who cares? After all, the experience of cognitive difficulty and the achievements enabled by it are just some prudential goods among many, and anyway may also be available outside cognitive work.

For children, the risk is moral and much more serious. The executive functions underwriting autonomous agency – capacities such as working memory, inhibitory control, planning, and cognitive flexibility – do not

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

mature automatically. They develop through repeated, deliberate encounters with cognitive tasks that load these capacities. Ubiquitous cognitive assistance threatens to deprive children of exactly these encounters, simultaneously reducing opportunities for productive struggle and undermining their incentives to engage in whatever struggle remains.

A world with ubiquitous cognitive assistance likely means that, at least on the margins, everyone – kids *and* adults – can, if they want, get away with being *more* cognitively lazy. Given our default incentives, I think that's probably not great.[18]

5 Further Issues

5.1 Introduction

The preceding sections defended three primary claims. First, the status quo for cognitive work is untenable. Second, the widespread displacement of humans from cognitive work may pose no special threat to human flourishing. Third, a world characterized by ubiquitous cognitive assistance carries distinctive moral risks we've largely ignored.

How might these claims go wrong? Four objections seem most serious. I'll take them in ascending order of threat, say briefly why each ultimately doesn't undermine the core argument, and then close.

5.2 The Framing

Labor economists may object that Section 2's application of the task-based framework oversimplifies the dynamics governing automation's effects on labor markets. The framework classically treats tasks as discrete, separable units, but cognitive work often involves *task bundles* whose value depends on integration rather than aggregation. More fundamentally, the equilibrium effects I dismissed as likely insufficient may prove more robust at preserving cognitive work than my analysis suggests. The historical record certainly supports cautious optimism about labor markets' capacity to absorb technological shocks, even dramatic ones (Autor 2015). And the mechanisms most likely to preserve cognitive work specifically – new cognitive tasks emerging faster than AI can perform them, or demand for cognitive output expanding enough to absorb productivity gains – aren't crazy. As I think of things, the deciding questions are *how capable* artificial systems get and *how cheap*. Equilibrium effects might

[18] Peter Richerson and Robert Boyd once quipped that we humans are "under stringent selection pressure to be as stupid as [we] can get away with" (Richerson and Boyd 2008).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

preserve cognitive work if capabilities plateau or costs stay high. But I know of no serious analysis suggesting business as usual in cognitive labor markets if we get human-level AI at a dollar a day.

These are empirical questions. But even if the economics for human cognitive workers proves more favorable than I've suggested, the normative analyses of Sections 3 and 4 still hold. Displacement's being less widespread doesn't make it threatening if it occurs; assistance's being less ubiquitous doesn't eliminate developmental risks where it appears. Getting the economics wrong would change the scale of the moral problems I've identified, not their character.

5.3 The Values

Philosophers of work may object that my analysis of what makes work good for human flourishing leaves out something important. While I've cataloged generic welfare goods associated with work (engagement, achievement, autonomy, and the rest), I've said little about the sense that one's work *matters* in a deep sense, for instance, by expressing or constituting one's identity (Bradford 2022; Moriarty 2009). Perhaps cognitive work provides privileged access to the good of meaningful work in ways that not-cognitive work doesn't. I'm uncertain how much weight to put on this idea, partly because "meaning" in this context remains frustratingly squishy as a concept: capacious enough to cover almost anything someone cares about, yet specific enough to sound like a determinate good. Rosso et al. (2010) provides an exhaustive overview of the range of things people mean when they say that work has (or should have) "meaning": I count more than 100 senses in that review. Moreover, whether meaning is a distinct welfare good is itself contested (Wolf 2010).

Epistemologists might press a sharper version of this worry. Scientific research and philosophical inquiry are arguably epistemic enterprises whose value depends partly on how the cognitive work they involve is structured and motivated (Kitcher 2001). If what matters isn't just *whether* the tasks get done but *who does them* and *to what end*, then displacement from these enterprises involves losses the parity claim can't capture. It's true that these activities have, until now, been the exclusive domain of human cognition. But it's unclear whether the epistemic goods themselves depend on human involvement or whether we've simply attached special significance to activities that have for all of history happened to require humans. In worlds where artificial systems perform the cognitive tasks

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

that largely comprise these enterprises, we may be left with something like an epiphenomenon: A halo of importance that traces to historical contingency rather than anything essential.

Even if I'm wrong about this – even if there are goods unique to human cognitive work that are genuinely lost in displacement – notice how narrow the resulting objection is. It targets the parity claim in Section 3, not the argument about assistance in Section 4. And it establishes at most that *some* displacement is unfortunate, not that it's the grave threat conventional wisdom often assumes.

5.4 The Developmental Evidence

Section 4's moral argument depends on an empirical claim: that executive functions develop through regular encounters with cognitive difficulty and that ubiquitous assistance will deprive children of these encounters.

Developmental psychologists may find the claim both oversimplified and underspecified. The relationship between challenge and development isn't linear; excessive difficulty impairs rather than builds capacity. The "appropriately challenging" engagement I invoked requires calibration that varies by individual and context. Perhaps well-designed AI assistance could provide precisely this calibration, adjusting difficulty dynamically in ways that human-structured environments never could. This is the optimistic vision of tech-ed enthusiasts and entrepreneurs such as Sal Khan (Khan 2024).

I welcome high-quality empirical evidence showing the impact of ubiquitous cognitive assistance on long-term cognitive outcomes for children. For reasons I've explained, I expect such evidence to show that ubiquitous cognitive assistance frustrates, rather than promotes, cognitive development. Importantly, empirical evidence that weakens the case I've made here would have to go further than showing that difficulty-*preserving* cognitive assistance (e.g., in the form of carefully designed AI tutors) is good for cognitive development. The relevant question here is what happens when assistance is ubiquitous, available by default, and used by children with the predictable aim of avoiding the aversive experience of cognitive effort.

5.5 Extensions

Two issues aren't objections to the core argument but suggest it extends further than I've been able to pursue here.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

First, consider political governance. Stable collective governance in most political arrangements requires a citizenry with robust cognitive capacities that look suspiciously like the executive functions Section 4 argued are threatened by ubiquitous assistance (cf. Barry 1995; Fumagalli 2023). If those capacities are affected population-wide, the result may be a polity that can no longer deliberate in the way democratic governance requires or behave stably in response to incentive structures in the way more technocratic systems assume. How an increasingly cognitively enfeebled citizenry might differentially affect stability under various forms of governance is a challenging question. The normative literature is characterized by a fraught debate that pits democrats (cf. Estlund 2009) against both epistocrats (cf. Brennan 2011) and Confucian meritocrats cf. Bell (2015). I think alternative models of democracy, such as government by sortition (cf. Guerrero 2024), also merit further attention in this context.

Second, consider developing economies. The argument throughout has focused on developed economies where cognitive work already dominates. But AI's impact will unfold globally, and the dynamics in developing economies may differ substantially. Development economists have raised concerns about premature deindustrialization: economies losing manufacturing jobs before reaching the income levels at which wealthy nations made the same transition (Rodrik 2016). A parallel worry is premature *de-cognitivization*: economies transitioning from manufacturing to services may find the cognitive work they were moving toward automated before they arrive. This, in turn, may point to increasing obligations of the global rich to the global poor. Unfortunately, we're arguably already failing to meet our existing obligations on that front.

5.6 The Real Danger Might Lie Elsewhere

AI safety researchers may find the entire framing of this Element parochial. I've assumed AI systems remain, fundamentally, more or less *really awesome* tools that broadly serve human interests (though perhaps not the interests of particular individuals). But the same rapid improvements in capabilities that make displacement likely might simultaneously make something else likely, too: AI systems that relentlessly pursue goals misaligned with human interests and that humanity as a whole can no longer control (cf. Yudkowsky and Soares 2025; Bostrom 2014; Hendrycks et al. 2023).

On this view, an academic voicing concerns about cognitive workers' wages and children's cognitive development in 2026 is a bit like

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Homo erectus grunting concerns about *Homo sapiens*' campfire smoke: They're about to have *much more serious* sapiens-related problems on their collective hands!

Perhaps. But existential risks, however large, remain deeply uncertain in both likelihood and tractability (Ord 2020; Sharadin 2025; Thorstad 2023). The transformation of cognitive work, meanwhile, is happening now, to people who already exist.

This Element has argued that conventional wisdom about AI and cognitive work gets the normative valence backward. Displacement sounds alarming but may be benign or even beneficial: Whatever makes work good for people is equally available outside cognitive work, and displacing humans from cognitive labor enables leveraged gains in research and access to justice. Assistance sounds gentle but carries a risk we have largely ignored: The systematic reduction of cognitive difficulty threatens the developmental conditions under which children acquire the capacities constitutive of autonomous agency. The most urgent practical implication is not about labor markets at all. It is that a technology currently being built and deployed at enormous scale, with essentially no consideration of developmental consequences, is poised to reshape the cognitive environment in which the next generation of human beings learns to think. Nobody has to be complicit in this. But complicity is the default, and defaults, as we know, tend to stick.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, vol. 4. Elsevier. https://doi.org/10.1016/S0169-7218(11)02410-5.

Acemoglu, Daron, and Pascual Restrepo. 2017. "Robots and Jobs: Evidence from US Labor Markets." Working Paper No. 23285. Working Paper Series. National Bureau of Economic Research, March. https://doi.org/10.3386/w23285.

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 108 (6): 1488–542. https://doi.org/10.1257/aer.20160696.

Acemoglu, Daron, and Pascual Restrepo. 2019. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." *Journal of Economic Perspectives* 33 (2): 3–30. https://doi.org/10.1257/jep.33.2.3.

Acemoglu, Daron, and Pascual Restrepo. 2020a. "The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand." *Cambridge Journal of Regions, Economy and Society* 13 (1): 25–35.

Acemoglu, Daron, and Pascual Restrepo. 2020b. "Unpacking Skill Bias: Automation and New Tasks." *AEA Papers and Proceedings* 110 (May): 356–61. https://doi.org/10.1257/pandp.20201063.

Adolph, Karen E., and John M. Franchak. 2017. "The Development of Motor Behavior." *Wiley Interdisciplinary Reviews: Cognitive Science* 8 (1–2): e1430.

Agrawal, Ajay K., Joshua S. Gans, and Avi Goldfarb. 2025. "The Economics of Bicycles for the Mind." Working Paper No. 34034. Working Paper Series. National Bureau of Economic Research, July. https://doi.org/10.3386/w34034.

Ahmed, Sammy F., Sandra Tang, Nicholas E. Waters, and Pamela Davis-Kean. 2019. "Executive Function and Academic Achievement: Longitudinal Relations from Early Childhood to Adolescence." *Journal of Educational Psychology* 111 (3): 446.

Alcott, Blake. 2005. "Jevons' Paradox." *Ecological Economics* 54 (1): 9–21.

Amodei, Dario. 2024. *Machines of Loving Grace*. October 11. https://darioamodei.com/machines-of-loving-grace.

Anderson, Elizabeth. 2017. *Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)*. Princeton University Press.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Anthropic. 2024. *The Claude 3 Model Family: Opus, Sonnet, Haiku*. https://docs.anthropic.com/en/docs/resources/claude-3-model-card.

Anthropic. 2025. *Effective Harnesses for Long-Running Agents*. www.anthropic.com/engineering/effective-harnesses-for-long-running-agents.

Applebaum, Herbert. 1992. *The Concept of Work: Ancient, Medieval, and Modern*. State University of New York Press.

Aristotle. 2014. *Aristotle: Nicomachean Ethics*. Cambridge University Press.

Arneson, Richard J. 1989. "Philosophical Studies." *An International Journal for Philosophy in the Analytic Tradition*. 56 (1): 77–93.

Arrow, Kenneth J. 1973. "Higher Education as a Filter." *Journal of Public Economics* 2 (3): 193–216.

Autor, David. 2022. *The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty*. National Bureau of Economic Research.

Autor, David. 2024. "Applying AI to Rebuild Middle Class Jobs." Working Paper No. 32140. Working Paper Series. National Bureau of Economic Research, February. https://doi.org/10.3386/w32140.

Autor, David, and Anna Salomons. 2018. *Is Automation Labor-Displacing? Productivity Growth, Employment, and the Labor Share*. National Bureau of Economic Research.

Autor, David H. 2014. "Skills, Education, and the Rise of Earnings Inequality Among the 'Other 99 Percent.'" *Science* 344 (6186): 843–51.

Autor, David H. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives* 29 (3): 3–30. https://doi.org/10.1257/jep.29.3.3.

Autor, David H., David Dorn, and Gordon H. Hanson. 2016. "The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade." *Annual Review of Economics* 8 (1): 205–40.

Autor, David H., and Michael J. Handel. 2013. "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages." *Journal of Labor Economics* 31 (2): S59–S96. https://doi.org/10.1086/669332.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in US Wage Inequality: Revising the Revisionists." *The Review of Economics and Statistics* 90 (2): 300–323.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118 (4): 1279–333.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Baddeley, Alan. 2007. *Working Memory, Thought, and Action*. Vol. 45. OUP Oxford.

Baddeley, Alan. 2020. "Working Memory." *Memory*, 71–111.

Baddeley, Alan D., and Graham James Hitch. 1974. "Working Memory (Vol. 8)." *In Recent Advances in Learning and Motivation*, edited by G. A. Bower, 60452–51.

Bai, Tongdong. 2019. *Against Political Equality: The Confucian Case*. Princeton University Press.

Barry, Brian. 1995. "John Rawls and the Search for Stability." University of Chicago Press.

Baumol, William. 2012. *The Cost Disease*. Yale University Press.

Baumol, William J. 1967. "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." *The American Economic Review* 57 (3): 415–26.

Bayern, Shawn. 2016. "The Implications of Modern Business–Entity Law for the Regulation of Autonomous Systems." *European Journal of Risk Regulation* 7 (2): 297–309.

Becker, Gary. 1994. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. University of Chicago Press.

Beckett, Celia, Barbara Maughan, Michael Rutter, et al. 2006. "Do the Effects of Early Severe Deprivation on Cognition Persist into Early Adolescence? Findings from the English and Romanian Adoptees Study." *Child Development* 77 (3): 696–711.

Bell, Daniel A. 2015. *The China Model: Political Meritocracy and the Limits of Democracy*. Princeton University Press.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23.

Berki, Borbála. 2022. "Overview of the Relationship between Human Spatial Abilities and GPS Usage." *2022 IEEE 1st International Conference on Cognitive Mobility (CogMob)*, 000019–22.

Bermúdez, Juan Pablo. 2025. "What Is the Feeling of Effort About?" *Australasian Journal of Philosophy* 103 (1): 88–105. https://doi.org/10.1080/00048402.2024.2351208.

Best, John R., and Patricia H. Miller. 2010. "A Developmental Perspective on Executive Function." *Child Development* 81 (6): 1641–60. https://doi.org/10.1111/j.1467-8624.2010.01499.x.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Best, John R., Patricia H. Miller, and Lara L. Jones. 2009. "Executive Functions after Age 5: Changes and Correlates." *Developmental Review: DR* 29 (3): 180–200. https://doi.org/10.1016/j.dr.2009.05.002.

Blinder, Alan S. 2006. "Offshoring: The Next Industrial Revolution?" *Foreign Affairs* 85 (2): 113–28. https://doi.org/10.2307/20031915.

Blinder, Alan S., and Alan B. Krueger. 2013. "Alternative Measures of Offshorability: A Survey Approach." *Journal of Labor Economics* 31 (S1): S97–S128.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb. 2020. "Are Ideas Getting Harder to Find?" *American Economic Review* 110 (4): 1104–44.

Boden, Margaret A. 2014. "GOFAI." *The Cambridge Handbook of Artificial Intelligence*, 89.

Boeken, Tom, Jean Feydy, Augustin Lecler, et al. 2023. "Artificial Intelligence in Diagnostic and Interventional Radiology: Where Are We Now?" *Diagnostic and Interventional Imaging* 104 (1): 1–5.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, et al. 2022. "On the Opportunities and Risks of Foundation Models." arXiv:2108.07258. Preprint, arXiv, July 12. https://doi.org/10.48550/arXiv.2108.07258.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Bovens, Luc. 2009. "The Ethics of Nudge." In *Preference Change: Approaches from Philosophy, Economics and Psychology*, edited by Till Grüne-Yanoff and Sven Ove Hansson. Springer Netherlands. https://doi.org/10.1007/978-90-481-2593-7_10.

Bradford, Gwen. 2015. *Achievement*. Oxford University Press.

Bradford, Gwen. 2022. "Achievement and Meaning in Life." In *The Oxford Handbook of Meaning in Life*, edited by Iddo Landau. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190063504.013.4.

Braem, Senne, and Tobias Egner. 2018. "Getting a Grip on Cognitive Flexibility." *Current Directions in Psychological Science* 27 (6): 470–76. https://doi.org/10.1177/0963721418787475.

Brennan, Jason. 2011. "The Right to a Competent Electorate." *The Philosophical Quarterly* 61 (245): 700–24.

Brennan, Jason. 2016. *Against Democracy*. Princeton University Press.

Brennan, Samantha, and Robert Noggle. 1997. "The Moral Status of Children: Children's Rights, Parents' Rights, and Family Justice." *Social Theory and Practice* 23 (1): 1–26.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Brighouse, Harry, and Adam Swift. 2014. *Family Values: The Ethics of Parent-Child Relationships*. Princeton University Press.

Browning, Jacob, and Yann LeCun. 2023. "Language, Common Sense, and the Winograd Schema Challenge." *Artificial Intelligence* 325 (December): 104031. https://doi.org/10.1016/j.artint.2023.104031.

Brynjolfsson, Erik. 1993. "The Productivity Paradox of Information Technology." *Communications of the ACM* 36 (12): 66–77.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. WW Norton & Company.

Buchanan, James M. 1991. "Opportunity Cost." In *The World of Economics* (pp. 520–525). Palgrave Macmillan UK.

Bullock, Justin B., Yu-Che Chen, Johannes Himmelreich, et al., eds. 2023. *Oxford Handbook of AI Governance*. Oxford University Press.

Burns, Mary, Rebecca Winthrop, Natasha Luther, Emma Venetis, and Rida Karim. 2026. *A New Direction for Students in an AI World: Prosper, Prepare, Protect*. Brookings Institution. www.brookings.edu/articles/a-new-direction-for-students-in-an-ai-world-prosper-prepare-protect/.

Caplow, Theodore. 1954. *The Sociology of Work*. University of Minnesota Press

Cappelen, Herman, and Josh Dever. 2025. "Going Whole Hog: A Philosophical Defense of AI Cognition." arXiv:2504.13988. Preprint, arXiv, April 18. https://doi.org/10.48550/arXiv.2504.13988.

Carneiro, Pedro, James J. Heckman, and Edward J. Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101 (6): 2754–81.

Center on the Developing Child. 2011. "Building the Brain's 'Air Traffic Control' System: How Early Experiences Shape the Development of Executive Function." Working Paper. Harvard.

Chalmers, David, and Andy Clark. 1998. "The Extended Mind." *Analysis* 58 (1): 7–19.

Cholbi, Michael. 2018. "The Duty to Work." *Ethical Theory and Moral Practice* 21 (5): 1119–33.

Cholbi, Michael. 2023. "Philosophical Approaches to Work and Labor." In *The Stanford Encyclopedia of Philosophy*, Summer 2023, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/work-labor/.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Chollet, François. 2019. "On the Measure of Intelligence." arXiv Preprint arXiv:1911.01547.

Christian Science Monitor. 1986. "The Great Calculator Debate. Educators Disagree over Their Place in the Classroom." www.csmonitor.com/1986/0509/dcalcf.html.

Christiano, Paul, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. "Deep Reinforcement Learning from Human Preferences." arXiv.org, June 12. https://arxiv.org/abs/1706.03741v4.

Clark, Andy. 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press. https://global.oup.com/academic/product/natural-born-cyborgs-9780195177510?cc=us&lang=en&.

Clark, Andy. 2010. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.

Cockburn, Iain M., Rebecca Henderson, and Scott Stern. 2018. "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb (pp. 115–146). University of Chicago Press.

Crawford, Matthew B. 2009. *Shop Class as Soulcraft: An Inquiry into the Value of Work*. Penguin.

Csikzentimihalyi, Mihaly. 1975. "Beyond Boredom and Anxiety: Experiencing Flow in Work and Play." San Francisco/Washington/London.

Csikszentmihalyi, Mihaly, Sami Abuhamdeh, and Jeanne Nakamura. 2014. "Flow." In *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 227–238). Springer.

Csikszentmihalyi, Mihaly, and Mihaly Csikzentmihaly. 1990. *Flow: The Psychology of Optimal Experience*. Vol. 1990. Harper & Row.

Dahmani, Louisa, and Véronique D. Bohbot. 2020. "Habitual Use of GPS Negatively Impacts Spatial Memory During Self-Guided Navigation." *Scientific Reports* 10 (1): 6310. https://doi.org/10.1038/s41598-020-62877-0.

Danaher, John. 2017. "Will Life Be Worth Living in a World without Work? Technological Unemployment and the Meaning of Life." *Science and Engineering Ethics* 23 (1): 41–64.

Danaher, John. 2019. *Automation and Utopia: Human Flourishing in a World without Work*. Harvard University Press.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

David, Louise, Eliana Vassena, and Erik Bijleveld. 2024. "The Unpleasantness of Thinking: A Meta-Analytic Review of the Association between Mental Effort and Negative Affect." *Psychological Bulletin* (US) 150 (9): 1070–93. https://doi.org/10.1037/bul0000443.

Davidson, Tom, Rose Hadshar, and William MacAskill. 2025. *Three Types of Intelligence Explosion*. Forethought. www.forethought.org/research/three-types-of-intelligence-explosion.

Deci, Edward L., and Richard M. Ryan. 2000. "The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior." *Psychological Inquiry* 11 (4): 227–68.

DeepSeek-AI, Daya Guo, Dejian Yang, et al. 2025. "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning." arXiv:2501.12948. Preprint, arXiv, January 22. https://doi.org/10.48550/arXiv.2501.12948.

Diamond, Adele. 2013. "Executive Functions." *Annual Review of Psychology* 64: 135–68. https://doi.org/10.1146/annurev-psych-113011-143750.

DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen. 2016. "Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs." *Journal of Health Economics* 47: 20–33.

Ding, Jeffrey. 2024. *Technology and the Rise of Great Powers: How Diffusion Shapes Economic Competition*. Princeton University Press

Doebel, Sabine. 2020. "Rethinking Executive Function and Its Development." *Perspectives on Psychological Science* 15 (4): 942–56. https://doi.org/10.1177/1745691620904771.

Dormann, Christian, and Dieter Zapf. 2001. "Job Satisfaction: A Meta-Analysis of Stabilities." *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 22 (5): 483–504.

Ellington, Aimee J. 2003. *A Meta-Analysis of the Effects of Calculators on Students' Achievement and Attitude Levels in Precollege Mathematics Classes*. Journal for Research in Mathematics Education. November 1. https://pubs.nctm.org/view/journals/jrme/34/5/article-p433.xml.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. "GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models." arXiv:2303.10130. Preprint, arXiv, August 21. https://doi.org/10.48550/arXiv.2303.10130.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Engel, Ernst. 1857. "Die Productions- Und Consumtionsverhältnisse Des Königreichs Sachsen." *Zeitschrift Des Statistischen Bureaus Des Königlich Sächsischen Ministeriums Des Innern,* nos. 8–9.

Ericsson, K. Anders, Ralf T. Krampe, and Clemens Tesch-Römer. 1993. "The Role of Deliberate Practice in the Acquisition of Expert Performance." *Psychological Review* 100 (3): 363.

Esping-Andersen, Gosta. 1990. *The Three Worlds of Welfare Capitalism*. Princeton University Press.

Estlund, David. 2009. *Democratic Authority: A Philosophical Framework*. Princeton University Press.

Evans, Jonathan St B. T. 2008. "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition." *Annual Review of Psychology* 59: 255–78. https://doi.org/10.1146/annurev.psych.59.103006.093629.

Feinberg, Joel. 1998. "The Child's Right to an Open Future." In *Philosophy of Education: Major Themes in the Analytic*, edited by Paul Hirst and Patricia White (pp. 250–270). Routledge

Felten, Edward W., Manav Raj, and Robert Seamans. 2018. "A Method to Link Advances in Artificial Intelligence to Occupational Abilities." *AEA Papers and Proceedings* 108: 54–57.

Felten, Edward, Manav Raj, and Robert Seamans. 2021. "Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses." *Strategic Management Journal* 42 (12): 2195–217.

Felten, Edward, Manav Raj, and Robert Seamans. 2023. "How Will Language Modelers Like ChatGPT Affect Occupations and Industries?" arXiv Preprint arXiv:2303.01157.

Fletcher, Guy. 2015. "Objective List Theories." In *The Routledge Handbook of Philosophy of Well-Being*, edited by Guy Fletcher (pp. 148–160). Routledge.

Franchak, John M., and Karen E. Adolph. 2024. "An Update of the Development of Motor Behavior." *Wiley Interdisciplinary Reviews: Cognitive Science* 15 (6): e1682.

Frankfurt, Harry. 2018. "Equality as a Moral Ideal." In *The Notion of Equality* (pp. 367–389). Routledge.

Franklin-Hall, Andrew. 2013. "On Becoming an Adult: Autonomy and the Moral Relevance of Life's Stages." *The Philosophical Quarterly* 63 (251): 223–47.

Frey, Carl Benedikt, and Michael A. Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114: 254–80.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Fumagalli, Corrado. 2023. "Stability in Liberal Epistocracies." *Social Epistemology*, January 2, 97–109. www.tandfonline.com/doi/abs/10.1080/02691728.2022.2131482.

Furendal, Markus, Huub Brouwer, and Willem van Der Deijl. 2024. "The Future of the Philosophy of Work." *Journal of Applied Philosophy* 41 (2): 181–201.

Fuster, Joaquín M. 2002. "Frontal Lobe and Cognitive Development." *Journal of Neurocytology* 31 (3): 373–85.

Gabriel, Iason, Arianna Manzini, Geoff Keeling, et al. 2024. "The Ethics of Advanced AI Assistants." arXiv:2404.16244. Preprint, arXiv, April 28. https://doi.org/10.48550/arXiv.2404.16244.

Gans, Joshua. 2026. "A Model of Artificial Jagged Intelligence." arXiv:2601.07573. Preprint, arXiv, January 12. https://doi.org/10.48550/arXiv.2601.07573.

Gans, Joshua S., and Avi Goldfarb. 2026. "O-Ring Automation." Working Paper No. 34639. Working Paper Series. National Bureau of Economic Research, January. https://doi.org/10.3386/w34639.

Gemini Team, Rohan Anil, Sebastian Borgeaud, et al. 2024. "Gemini: A Family of Highly Capable Multimodal Models." arXiv:2312.11805. Preprint, arXiv, April 2. https://doi.org/10.48550/arXiv.2312.11805.

Geroski, Paul A. 2000. "Models of Technology Diffusion." *Research Policy* 29 (4–5): 603–25.

Gheaus, Anca. 2021. "The Best Available Parent." *Ethics* 131 (3): 431–59.

Gheaus, Anca, and Lisa Herzog. 2016. "The Goods of Work (Other Than Money!)." *Journal of Social Philosophy* 47 (1): 70–89

Gheaus, Anca, Gideon Calder, and Jurgen De Wispelaere, eds. 2019. *The Routledge Handbook of the Philosophy of Childhood and Children* (pp. 90–100). Routledge.

Giesinger, Johannes. 2012. "Kant on Dignity and Education." *Educational Theory* 62 (6): 609–20.

Gigerenzer, Gerd, and Daniel G. Goldstein. 1996. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." *Psychological Review* (US) 103 (4): 650–69. https://doi.org/10.1037/0033-295X.103.4.650.

Gmyrek, Pawel, Janine Berg, and David Bescond. 2023. "Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality." ILO Working Paper 96.

Goldin, Claudia, and Lawrence F. Katz. 2007. "The Race between Education and Technology: The Evolution of US Educational Wage

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Differentials, 1890 to 2005." National Bureau of Economic Research Cambridge, MA., USA.

Google. 2025 "What Is Artificial General Intelligence?" Accessed August 15, 2025. https://cloud.google.com/discover/what-is-artificial-general-intelligence.

Goos, Maarten, Alan Manning, and Anna Salomons. 2014. "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring." *American Economic Review* 104 (8): 2509–26.

Gordon, Robert J. 2016. *The Rise and Fall of American Growth: The US Standard of Living since the Civil War*. Princeton University Press.

Gottweis, Juraj, Wei-Hung Weng, Alexander Daryin, et al. 2025. "Towards an AI Co-Scientist." arXiv:2502.18864. Preprint, arXiv, February 26. https://doi.org/10.48550/arXiv.2502.18864.

Gregory, Alex. 2015. "Hedonism." In *The Routledge Handbook of Philosophy of Wellbeing*, edited by Guy Fletcher (pp. 113–123). Routledge.

Grinschgl, Sandra, Frank Papenmeier, and Hauke S. Meyerhoff. 2021. "Consequences of Cognitive Offloading: Boosting Performance but Diminishing Memory." *Quarterly Journal of Experimental Psychology* 74 (9): 1477–96.

Guerrero, Alexander. 2024. *Lottocracy: Democracy Without Elections*. Oxford University Press.

Haidt, Jonathan. 2024. *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness*. Penguin.

Hajek, Alan. 2016. "Philosophical Heuristics and Philosophical Methodology." In *The Oxford Handbook of Philosophical Methodology*, edited by Herman Cappelen (pp. 348–373). Oxford University Press.

Hao, Karen. 2025. *Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI*. Penguin Group.

Harding, Jacqueline, and Nathaniel Sharadin. 2024. "What Is It for a Machine Learning System to Have a Capability?" *British Journal for the Philosophy of Science*.

Hayes, Chris. 2025. *The Sirens' Call: How Attention Became the World's Most Endangered Resource*. Random House.

Hayes, Joseph. 2017. "The Victorian Belief That a Train Ride Could Cause Instant Insanity." *Atlas Obscura, May*. www.atlasobscura.com/articles/railway-madness-victorian-trains.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Heathwood, Chris. 2015. "Desire-Fulfillment Theory." In *The Routledge Handbook of the Philosophy of Well-Being*, edited by Guy Fletcher (pp. 135–147). Routledge.

Hendrycks, Dan, Mantas Mazeika, and Thomas Woodside. 2023. "An Overview of Catastrophic AI Risks." arXiv Preprint arXiv:2306.12001.

Herd, Pamela, and Donald P. Moynihan. 2019. *Administrative Burden: Policymaking by Other Means*. Russell Sage Foundation.

Hirji, Sukaina. 2019. "Not Always Worth the Effort: Difficulty and the Value of Achievement." *Pacific Philosophical Quarterly* 100 (2): 525–48.

Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, et al. 2022. "Training Compute-Optimal Large Language Models." arXiv:2203.15556. Preprint, arXiv, March 29. https://doi.org/10.48550/arXiv.2203.15556.

Hoppe, Thomas, Deborah Schanz, Susann Sturm, and Caren Sureth-Sloane. 2023. "The Tax Complexity Index–a Survey-Based Country Measure of Tax Code and Framework Complexity." *European Accounting Review* 32 (2): 239–73.

Huang, Lei, Weijiang Yu, Weitao Ma, et al. 2025. "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions." *ACM Transactions on Information Systems* 43 (2): 1–55.

Humphrey, Stephen E., Jennifer D. Nahrgang, and Frederick P. Morgeson. 2007. "Integrating Motivational, Social, and Contextual Work Design Features: A Meta-Analytic Summary and Theoretical Extension of the Work Design Literature." *Journal of Applied Psychology* 92 (5): 1332–56. https://doi.org/10.1037/0021-9010.92.5.1332.

Hurka, Thomas. 2020. "The Parallel Goods of Knowledge and Achievement." *Erkenntnis* 85 (3): 589–608. https://doi.org/10.1007/s10670-020-00245-0.

Inzlicht, Michael, Amitai Shenhav, and Christopher Y. Olivola. 2018. "The Effort Paradox: Effort Is Both Costly and Valued." *Trends in Cognitive Sciences* 22 (4): 337–49.

Ionescu, Thea. 2012. "Exploring the Nature of Cognitive Flexibility." *New Ideas in Psychology* 30 (2): 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001.

Jachimowicz, Jon M., Shannon Duncan, Elke U. Weber, and Eric J. Johnson. 2019. "When and Why Defaults Influence Decisions: A

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Meta-Analysis of Default Effects." *Behavioural Public Policy* 3 (2): 159–86. https://doi.org/10.1017/bpp.2018.43.

James, Alexander, Timothy Retting, Jason F. Shogren, Brett Watson, and Samuel Wills. 2022. "Sovereign Wealth Funds in Theory and Practice." *Annual Review of Resource Economics* 14: 621–46. https://doi.org/10.1146/annurev-resource-111920-015758.

James, William. 1890. *The Principles of Psychology*. Edited by Frederick Burkhardt, Fredson Bowers, and Ignas K Skrupskelis. Vol. 1. Macmillan London.

Jaworska, Agnieszka, and Julie Tannenbaum. 2018. "The Moral Status of Children." In *The Routledge Handbook of the Philosophy of Childhood and Children* (pp. 67–78). Routledge.

Jones, Charles I. 2005. "Growth and Ideas." In *Handbook of Economic Growth*, vol. 1 (pp. 1063–1111). Elsevier.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kant, Immanuel. 1785. *Groundwork for the Metaphysics of Morals*. Oxford University Press.

Kaplan, Jared, Sam McCandlish, Tom Henighan, et al. 2020. "Scaling Laws for Neural Language Models." arXiv Preprint arXiv:2001.08361.

Keller, Simon. 2004. "Welfare and the Achievement of Goals." *Philosophical Studies* 121 (1): 27–41.

Khan, Sal. 2024. *Brave New Words: How AI Will Revolutionize Education (and Why That's a Good Thing)*. Penguin Publishing Group. https://books.google.com.sg/books?id=QMLQEAAAQBAJ.

Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford University Press.

Korinek, Anton, and Donghyun Suh. 2024. *Scenarios for the Transition to AGI*. National Bureau of Economic Research.

Kortum, Samuel S. 1997. "Research, Patenting, and Technological Change." *Econometrica* 65 (6): 1389–419.

Kremer, Michael. 1993. "The O-Ring Theory of Economic Development." *The Quarterly Journal of Economics* 108 (3): 551–75.

Kruse, Hagen, Emmanuel Mensah, Kunal Sen, and Gaaitzen De Vries. 2022. "A Manufacturing (Re) Naissance? Industrialization in the Developing World." *IMF Economic Review* 71 (2): 439.

Kulveit, Jan, Raymond Douglas, Nora Ammann, Deger Turan, David Krueger, and David Duvenaud. 2025. "Gradual

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Disempowerment: Systemic Existential Risks from Incremental AI Development." arXiv Preprint arXiv:2501.16946.

Kwa, Thomas, Ben West, Joel Becker, et al. 2025. "Measuring AI Ability to Complete Long Tasks." arXiv:2503.14499. Preprint, arXiv, March 30. https://doi.org/10.48550/arXiv.2503.14499.

Lake, Brenden M., Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. 2017. "Building Machines That Learn and Think Like People." *Behavioral and Brain Sciences* 40: e253.

Lei, Ya-Wen, and Rachel Kim. 2024. "Automation and Augmentation: Artificial Intelligence, Robots, and Work." *Annual Review of Sociology* 50 (1): 251–272.

Levinstein, Ben. 2023. "A Conceptual Guide to Transformers: Part I." Substack Newsletter. *The End of Reckoning*, January 31. https://benlevinstein.substack.com/p/a-conceptual-guide-to-transformers.

Levy, Neil. 2018. "Why Nudging Is No More Paternalistic Than Arguing." *The Philosophers' Magazine* 83: 53–59. https://doi.org/10.5840/tpm201883102.

Levy, Neil. 2019. "Nudge, Nudge, Wink, Wink: Nudging Is Giving Reasons." *Ergo: An Open Access Journal of Philosophy* 6. https://doi.org/10.3998/ergo.12405314.0006.010.

Lewis, David K. 1973. *Counterfactuals*. Blackwell.

Lobo, Michele A., and James C. Galloway. 2012. "Enhanced Handling and Positioning in Early Infancy Advances Development Throughout the First Year." *Child Development* 83 (4): 1290–302.

Lowrey, Annie. 2021. "How Government Learned to Waste Your Time." *The Atlantic*. www.theatlantic.com/politics/archive/2021/07/how-government-learned-waste-your-time-tax/619568/.

MacIntyre, Alasdair. 2007. *After Virtue: A Study in Moral Theory*. 3rd ed. University of Notre Dame Press.

Marcus, Gary. 2022. "Deep Learning Is Hitting a Wall." *Nautilus* 10: 2022.

Masley, Andy. 2025. "Empire of AI Is Wildly Misleading on AI Water Use." Substack newsletter. *Andy Masley*, November 17. https://andymasley.substack.com/p/empire-of-ai-is-wildly-misleading.

Matthias, Andreas. 2004. "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata." *Ethics and Information Technology* 6 (3): 175–83.

McCoy, Shelly S., Laura M. Dimler, and Luiza Rodrigues. 2025. "Parenting in Overdrive: A Meta-Analysis of Helicopter Parenting Across Multiple Indices of Emerging Adult Functioning." *Journal*

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

of Adult Development 32 (3): 222–45. https://doi.org/10.1007/s10804-024-09496-5.

Milgrom, Paul, and John Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *The American Economic Review* 80 (3): 511–28.

Miller, Claire Cain, and Sarah Mervosh. 2025. "How Much Screen Time Is Your Child Getting at School? We Asked 350 Teachers." *The Upshot*. *The New York Times*, November 12. www.nytimes.com/2025/11/12/upshot/teachers-survey-chromebooks-class.html.

Miyake, Akira, Naomi P. Friedman, Michael J. Emerson, Alexander H. Witzki, Amy How-erter, and Tor D. Wager. 2000. "The Unity and Diversity of Executive Functions and Their Contributions to Complex 'Frontal Lobe' Tasks: A Latent Variable Analysis." *Cognitive Psychology* 41 (1): 49–100.

Moilanen, Kristin L., and Mary Lynn Manuel. 2019. "Helicopter Parenting and Adjustment Outcomes in Young Adulthood: A Consideration of the Mediating Roles of Mastery and Self-Regulation." *Journal of Child and Family Studies* 28 (8): 2145–58.

Mokyr, Joel. 2011. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press.

Mollick, E. 2024. *Co-Intelligence: Living and Working with AI*. Penguin Publishing Group. https://books.google.com.sg/books?id=J2HgEAAAQBAJ.

Monsell, Stephen, and Jon Driver. 2000. *Control of Cognitive Processes: Attention and Performance XVIII*. Vol. 18. MIT Press.

Moravec, Hans. 1988. *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press.

Moriarty, Jeffrey. 2009. "Rawls, Self-Respect, and the Opportunity for Meaningful Work." *Social Theory and Practice* 35 (3): 441–59.

Munakata, Yuko, Seth A. Herd, Christopher H. Chatham, Brendan E. Depue, Marie T. Banich, and Randall C. O'Reilly. 2011. "A Unified Framework for Inhibitory Control." *Trends in Cognitive Sciences* 15 (10): 453–59.

Narayanan, Arvind, and Sayash Kapoor. 2025. *AI as Normal Technology*. Knight First Amendment Institute.

National Assessment of Educational Progress. 2022. *NAEP Mathematics: Mathematics Results*. National Assessment of Educational Progress. www.nationsreportcard.gov/mathematics/.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Nogrady, Bianca. 2024. "Australia Bans Under-16s from Social Media 'to Protect Their Development.'" *British Medical Journal* 387: q2724.

Nordhaus, William D. 2006. "Baumol's Diseases: A Macroeconomic Perspective." National Bureau of Economic Research Cambridge, MA, USA.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Basic Books.

OpenAI. 2022. "ChatGPT: Optimizing Language Models for Dialogue." November 30. https://openai.com/blog/chatgpt/.

OpenAI. 2024. "OpenAI Charter." https://openai.com/charter/.

Ord, Toby. 2020. *The Precipice: Existential Risk and the Future of Humanity*. Hachette UK.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford University Press.

Parfit, Derek. 2018. "Equality and Priority 1." In *The Notion of Equality*, edited by Mane Hajdin (pp. 427–446). Routledge.

Penadés, José R., Juraj Gottweis, Lingchen He, et al. 2025. "AI Mirrors Experimental Science to Uncover a Novel Mechanism of Gene Transfer Crucial to Bacterial Evolution." Preprint, bioRxiv, February 19. https://doi.org/10.1101/2025.02.19.639094.

Pettit, Philip. 1997. *Republicanism: A Theory of Freedom and Government*. Oxford University Press.

Pettit, Philip. 2012. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge University Press.

Phan, Long, Alice Gatti, Ziwen Han, et al. 2025. "Humanity's Last Exam." arXiv Preprint arXiv:2501.14249.

Pizzinelli, Carlo, Augustus J. Panton, Ms Marina Mendes Tavares, Mauro Cazzaniga, and Longji Li. 2023. *Labor Market Exposure to AI: Cross-Country Differences and Distributional Implications*. International Monetary Fund.

Plato. 1997. *Plato: Complete Works*. Edited by John M. Cooper and Douglas S. Hutchinson. Hackett Publishing.

Polanyi, Michael. 1966. *The Tacit Dimension*. Routledge & Kegan Paul.

Rawls, John. 1971. *A Theory of Justice: Original Edition*. Belknap Press.

Rawls, John. 2001. *Justice as Fairness: A Restatement*. Harvard University Press.

Rhode, Deborah L. 2004. *Access to Justice*. Oxford University Press.

Richerson, Peter J., and Robert Boyd. 2008. *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Risko, Evan F., and Sam J. Gilbert. 2016. "Cognitive Offloading." *Trends in Cognitive Sciences* 20 (9): 676–88.

Rodrik, Dani. 2016. "Premature Deindustrialization." *Journal of Economic Growth* 21 (1): 1–33.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5, *Part* 2): S71–S102.

Rosso, Brent D., Kathryn H. Dekas, and Amy Wrzesniewski. 2010. "On the Meaning of Work: A Theoretical Integration and Review." *Research in Organizational Behavior* 30: 91–127.

Roth, Klas, and Chris Surprenant. 2012. *Kant and Education: Interpretations and Commentary*. Routledge.

Ruffini, Costanza, Fatbardha Osmani, Chiara Martini, Winnie-Karen Giera, and Chiara Pecini. 2024. "The Relationship Between Executive Functions and Writing in Children: A Systematic Review." *Child Neuropsychology* 30 (1): 105–63.

Sandefur, Rebecca L. 2019. "Access to What?" *Daedalus* 148 (1): 49–55.

Schapiro, Tamar. 1999. "What Is a Child?" *Ethics* 109 (4): 715–38.

Schiffrin, Holly H., Miriam Liss, Haley Miles-McLean, Katherine A. Geary, Mindy J. Erchull, and Taryn Tashner. 2014. "Helping or Hovering? The Effects of Helicopter Parenting on College Students' Well-Being." *Journal of Child and Family Studies* 23 (3): 548–57.

Sennett, Richard. 2008. *The Craftsman*. Yale University Press.

Sevilla, Jaime, and Edu Roldán. 2024. "Training Compute of Frontier AI Models Grows by 4-5x Per Year." https://epoch.ai/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year.

Sharadin, Nathaniel. 2025. "Promotionalism, Orthogonality, and Instrumental Convergence." *Philosophical Studies* 182 (7): 1–31.

Sharadin, Nathaniel. 2026a. "Co-Opting Artificial Achievement." Unpublished manuscript.

Sharadin, Nathaniel. 2026b. "Loss Function." Unpublished manuscript.

Smith, Edward E., and John Jonides. 1999. "Storage and Executive Processes in the Frontal Lobes." *Science* 283 (5408): 1657–61.

Sparrow, Betsy, Jenny Liu, and Daniel M. Wegner. 2011. "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips." *Science* 333 (6043): 776–78.

Spence, Michael. 1978. "Job Market Signaling." In *Uncertainty in Economics* (pp. 281–306). Academic Press.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Stanford HAI. 2025. *The 2025 AI Index Report*. https://hai.stanford.edu/ai-index/2025-ai-index-report.

Stanovich, Keith E. 2020. "Why Humans Are Cognitive Misers and What It Means for the Great Rationality Debate." In *Routledge Handbook of Bounded Rationality*, edited by Riccardo Viale (pp. 196–206). Routledge.

Sunstein, Cass R. 2021. *Sludge: What Stops Us from Getting Things Done and What to Do about It*. MIT Press.

Sunstein, Cass R. 2022. "Sludge Audits." *Behavioural Public Policy* 6 (4): 654–73.

Székely, Marcell, and John Michael. 2021. "The Sense of Effort: A Cost-Benefit Theory of the Phenomenology of Mental Effort." *Review of Philosophy and Psychology* 12 (4): 889–904.

Tan, Charlene. 2017. "Confucianism and Education." In *Oxford Research Encyclopedia of Education*, edited by Kathy Hytten.

Thaler, Richard H., and Cass R. Sunstein. 2021. *Nudge: The Final Edition*. Yale University Press.

Thorstad, David. 2023. "High Risk, Low Reward: A Challenge to the Astronomical Value of Existential Risk Mitigation." *Philosophy & Public Affairs* 51 (4): 373–412.

Thorstad, David. 2024. "Against the Singularity Hypothesis." *Philosophical Studies*, 1–25.

Trout, Cristian. 12/17/2025 7:52:57 PM. "Healthy Insurance Markets Will Be Critical for AI Governance." *Lawfare*, 12/17/2025 7:52:57 PM. www.lawfaremedia.org/article/healthy-insurance-markets-will-be-critical-for-ai-governance.

Trump, Donald. 2025. *Advancing Artificial Intelligence Education for American Youth. Executive Order*. www.whitehouse.gov/presidential-actions/2025/04/advancing-artificial-intelligence-education-for-american-youth/.

Uncapher, Melina R., Monica K. Thieu, and Anthony D. Wagner. 2016. "Media Multitasking and Memory: Differences in Working Memory and Long-Term Memory." *Psychonomic Bulletin & Review* 23: 483–90.

UNICEF Australia. n.d. "Social Media Ban in Australia – A Simple Guide." UNICEF Australia. Accessed January 21, 2026. www.unicef.org.au/unicef-youth/staying-safe-online/social-media-ban-explainer.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Van Parijs, Philippe, and Yannick Vanderborght. 2017. *Basic Income: A Radical Proposal for a Free Society and a Sane Economy*. Harvard University Press.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, et al. 2017. "Attention Is All You Need." *Advances in Neural Information Processing Systems* 30.

Villalobos, Pablo, Jaime Sevilla, Lennart Heim, Tamay Besiroglu, Marius Hobbhahn, and Anson Ho. 2022. "Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning." arXiv Preprint arXiv:2211.04325 1: 1.

Wahlers, Justin N., Katie E. Garrison, and William P. Hart. 2025. "Measuring Subjective Mental Effort as an Aversive and Enjoyable Phenomenon." *Motivation Science* (US) 11 (4): 425–38. https://doi.org/10.1037/mot0000393.

Walberg, Herbert J. 1984. "Improving the Productivity of America's Schools." *Educational Leadership* 41 (8): 19–27.

Wason, Peter C., and J. St B. T. Evans. 1974. "Dual Processes in Reasoning?" *Cognition* 3 (2): 141–54.

Watson, Joseph L., David Juergens, Nathaniel R. Bennett, et al. 2023. "De Novo Design of Protein Structure and Function with RFdiffusion." *Nature* 620 (7976): 1089–100.

Watzl, Sebastian. 2011. "The Philosophical Significance of Attention." *Philosophy Compass* 6 (10): 722–33. https://doi.org/10.1111/j.1747-9991.2011.00432.x.

Webb, Michael. 2019. "The Impact of Artificial Intelligence on the Labor Market." Available at SSRN 3482150.

Wei, Jason, Xuezhi Wang, Dale Schuurmans, et al. 2022. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." *Advances in Neural Information Processing Systems* 35: 24824–37.

Wolf, Susan. 2010. *Meaning in Life and Why It Matters*. Princeton University Press.

Yao, Shunyu, Dian Yu, Jeffrey Zhao, et al. 2023. "Tree of Thoughts: Deliberate Problem Solving with Large Language Models." *Advances in Neural Information Processing Systems* 36: 11809–22.

Yeung, Nick, and Christopher Summerfield. 2012. "Metacognition in Human Decision-Making: Confidence and Error Monitoring." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1594): 1310–21. https://doi.org/10.1098/rstb.2011.0416.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Yudkowsky, Eliezer, and Nate Soares. 2025. *If Anyone Builds It, Everyone Dies: Why Superhuman AI Would Kill Us All*. Hachette Publishing Group.

Zelazo, Philip David, Fergus I. M. Craik, and Laura Booth. 2004. "Executive Function Across the Life Span." *Acta Psychologica, Executive Control of Human Action*, 115 (2): 167–83. https://doi.org/10.1016/j.actpsy.2003.12.005.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Acknowledgments

For helpful feedback on earlier versions of this material, I'm grateful to audiences and seminar participants at the Columbia River Ethics Workshop, the Havana Workshop in Normative Philosophy, Lingnan University, the National University of Singapore, Notre Dame's Institute for Ethics and the Common Good (né Institute for Advanced Studies), NYU Abu Dhabi, the Royal Institute of Philosophy, South China Normal University, the University of Hong Kong, the University of Pennsylvania, and to members of the Cottage Group and the Hong Kong Catastrophic Risk Center Reading Group. Jana Clever listened to more about AI and cognitive work than anyone should have to.

In the course of researching this Element, the author used AI assistance of the kind it describes.

This work was supported by funding from the Hong Kong Research Grants Council (Grant number: 17604824).

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Cambridge Elements

Philosophy and AI

Herman Cappelen
University of Hong Kong

Herman Cappelen is a Chair Professor at the University of Hong Kong and the founder and director of the AI&Humanity-Lab. He has worked in many areas of philosophy, including philosophy of language, conceptual engineering, meta-philosophy, and the philosophy of AI. His most influential works include *Fixing Language: An Essay on Conceptual Engineering* (OUP, 2018), *Philosophy without Intuitions* (OUP, 2012), *The Concept of Democracy* (OUP, 2023), *Making AI Intelligible* (OUP, 2021), *Relativism and Monadic Truth* (OUP, 2009, with John Hawthorne), *Insensitive Semantics* (Blackwell, 2005 with Ernie Lepore), *The Inessential Indexical* (OUP, 2013, with Josh Dever). Cappelen and Josh Dever have written a three-volume introduction to contemporary philosophy of language: *Context and Communication* (OUP, 2016), *Puzzles of Reference* (OUP, 2018), and *Bad Language* (OUP, 2019). Cappelen has previously been a Chair Professor at The University of St Andrews, an associate professor at Oxford University, and a professor at The University of Oslo.

About the Series

The burning question of our time is what it means to be human in the age of Artificial Intelligence. Philosophy plays a unique and crucial role in answering this question. Philosophers investigate whether AI can understand language, be creative, possess emotions, be conscious, have intentions, make plans, execute intentional actions, be moral or immoral, empathize, possess knowledge, and exhibit wisdom. The answers to these and many related questions not only are intellectually challenging but also have potential real-world implications for how AI is developed and used.

Cambridge Elements in Philosophy and AI explores these questions and their ramifications. Each Element provides a survey of the literature, which will be a reliable resource for researchers and students, and also develops new ideas and arguments from the author's viewpoint. The authors of the Elements include some of the most prominent senior figures and up-and-coming junior scholars in the field.

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press

Cambridge Elements

Philosophy and AI

Elements in the Series

Emerging Questions in AI Welfare
Geoff Keeling and Winnie Street

AI and Consciousness: A Skeptical Overview
Eric Schwitzgebel

Geopolitics of Artificial Intelligence
Boris Babic and Brian Wong Yue Shun

AI & Cognitive Work
Nathaniel P. Sharadin

A full series listing is available at: www.cambridge.org/EPAI

https://doi.org/10.1017/9781009766081 Published online by Cambridge University Press