The Clock, Not the Culture

Predicting Workplace Behavior from the Individual to the Multinational in the Age of AI Colleagues

Ioan-Codrut Lazaroiu

Working paper, version 1.6, September 2026

Codrut Lazaroiu Advisory, Vienna. Correspondence: codrut@lazaroiu.at. Prepared during the MIT Sloan Executive Program in General Management (Cohort 15) and the MIT Sloan and Emeritus Negotiation and Influence programme.

Statement on the use of AI. Literature search, drafting and figure production were assisted by an AI system (Anthropic's Claude) working under the author's direction, and the same system executed the empirical study of Chapter 7a — building the harness, administering the instruments over the public APIs, and scoring the results — under the author's direction and with a model from its own family among the subjects, a conflict disclosed in that chapter. The research question, framework, study design decisions, field observations, questionnaire answers, judgements and conclusions are the author's, who takes full responsibility for the content.

Abstract

Behavioral assessment predicts individual workplace behavior, but weakly and conditionally: personality explains a small share of performance variance and only where the situation leaves room for it. This paper asks whether prediction survives the move to higher levels, from the individual to the human team, to the team that includes artificial-intelligence colleagues, to the firm in one country, and to the firm operating across several. It argues that each transition introduces a distinct translation problem (emergence, trust calibration, structure and leadership, cultural moderation, ecosystem fit) and specifies the rule that governs each. To make the argument testable, it maps the framework onto a single, fully measured evolutionary case, the medium ground finch of Daphne Major under drought, abundance, competitor arrival and hybridisation, and treats the arrival of AI in the workplace as a disturbance of the same structural kind. The breeder's equation is proposed as a first-order estimator of organisational response. The framework is then set against nine countries in which the author held senior management roles, using a structured questionnaire and the author's own assessment profile as a disclosed instrument. At the human-AI level the paper contributes original data: an empirical study administering a DISC-format instrument, a Big Five inventory and a behavioral game battery to AI colleagues configured as real functions configure them, across four model

versions from two providers (8,960 measurement units). The function that configures the colleague explains 65 to 92 per cent of the variance in its measured style and the model behind it less than 10 per cent, while behavior under provocation belongs to the model family and shifts with version updates the questionnaire cannot see. Thirteen propositions follow, two now carrying measurements. Two field findings are new: the role assigned to an advisor, human or artificial, tracks the organisation's decision latency rather than its national culture; and the predictability of workplace behavior depends less on how strong the situation is than on whether the source of its strength is codified. The paper closes with what the framework implies for practice and with the limits of a single-observer field lens.

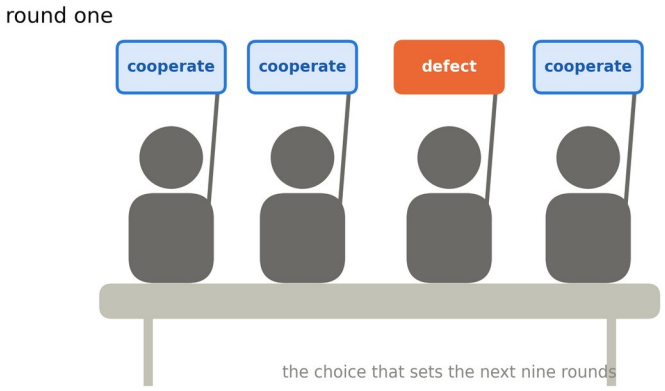
Acknowledgements

The paper was written during the MIT Sloan Executive Program in General Management (Cohort 15) and draws on the teaching of Professor Charles H. Fine, whose clockspeed framework and China-speed working paper anchor Chapters 4 and 8, and of Professor David Robertson, Professor Erin Scott and Dr Nick van der Meulen, whose EPGM sessions supplied the firm-level cases. It also draws on the MIT Sloan and Emeritus Negotiation and Influence programme, whose Module 2 exercise opens Chapter 1. None of them has reviewed the paper, and its errors and opinions are the author's alone.

1. Introduction

1.1 Round one

In a four-party exercise used in negotiation teaching, each party chooses in every round whether to cooperate or defect, and the payoffs are structured so that unanimous cooperation yields the highest joint total while unilateral defection yields the highest individual one. Across cohorts, the round-one choice predicts the whole trajectory: groups that begin with four cooperators tend to sustain cooperation through the bonus rounds and finish with the maximum score, and groups that begin with a defector rarely recover. In the cohort data reviewed for this paper, one group scored 100, the theoretical maximum, and another scored minus four (Emeritus, 2026). The exercise is a metaphor for the prisoner's dilemma, but it is also a small experiment in behavioral prediction. Four people, one observable choice each, and the outcome of the next nine rounds is largely set.



Round one: four parties, one visible choice each — and the trajectory of the next nine rounds is largely set (concept panel; the exercise of Section 1.1).

If a single early choice by four strangers predicts a group's fate over an hour, what predicts the behavior of a project team over a year, of a team that now includes a machine that gives advice, of a company under a new competitor, or of that company entering a country it has never operated in? Each of these is a larger version of the same question, and each is usually answered by a different discipline with a different instrument and a different idea of what counts as evidence. This paper asks the question once, across all of them.

1.2 The question

The research question is: at which levels of organisation, and under what conditions, does workplace behavior become predictable from measured characteristics, and what is lost at each transition from a lower level to a higher one?

The question presupposes that something is predictable at the lowest level, and the evidence supports that presupposition with a caveat. Validated personality instruments predict individual job performance at correlations between 0.2 and 0.3 for the best single trait (Sackett, Zhang, Berry & Lievens, 2022), and the popular practitioner instruments predict the form of behavior more reliably than its outcome (Chapter 5). The presupposition that anything is predictable at the highest level is more doubtful. National culture scores predict individual attitudes at about 0.18 (Taras, Kirkman & Steel, 2010) and firm behavior less well, and the mechanisms by which a firm adapts to a foreign environment are described in the literature mainly as costs and frictions (Zaheer, 1995; Kostova & Roth, 2002) rather than as predictable responses.

Between the two ends sit the team and, newly, the team with artificial members. Team composition research finds that member personality predicts team performance weakly and that the choice of aggregation rule matters more than the trait (Bell, 2007; Peeters, van Tuijl, Rutte & Reymen, 2006). Human-AI teaming research finds that combinations of people and models on average perform worse than the better of the two alone (Vaccaro, Almaatouq & Malone, 2024), which makes the human-AI team the level at which the paper's question is most open.

1.3 Why now

Three developments make the question timely. The first is the arrival of AI colleagues. Large language models entered daily knowledge work in 2023 and, by the time of writing, are embedded in the office software of most large organisations; the European Union's transparency obligations for AI systems that interact with people came into force on 2 August 2026 (Regulation (EU) 2024/1689, Article 50). A team with an AI member is no longer a laboratory condition. The second is the evidence that the effect of AI on work is uneven in a way that resembles selection: it raises the performance of some workers on some tasks and lowers it on others, along a frontier that shifts with each model release (Dell'Acqua et al., 2026; Brynjolfsson, Li & Raymond, 2025). The third is the argument, made most sharply by Fine (2026), that organisations whose internal clock runs slower than their industry's are selected out, and that AI has accelerated the industry clock in nearly every sector at once. If that is right, the question of how organisations respond behaviorally to a disturbance is not academic.

1.4 The approach

The paper is conceptual, with an illustrative field application and one original empirical study. It builds a five-level framework (individual, human team, human-AI team, firm, multinational), states the translation rule at each boundary, and derives thirteen propositions. To keep the framework honest about mechanism, it maps every level onto a single biological case in which selection, heritability and response have all been measured in the wild: the medium ground finch on Daphne Major over forty years (Grant & Grant, 2014). The finch chapter is not an ornament; it supplies the estimation tool (the breeder's equation) and the discipline of grading each analogy by how much mechanism it actually shares.

At the human-AI level, where the literature is thinnest and the paper's claims would otherwise rest entirely on other people's experiments, the paper runs its own: Chapter 7a administers the instruments of Chapter 5 and a behavioral battery to AI colleagues configured with the roles, documents and task regimes of six corporate functions, across four model versions from two providers, and tests two of the propositions directly.

The field application uses nine countries in which the author held senior management roles between 2006 and 2026, all in energy infrastructure. The observations were collected with a structured questionnaire aligned to the five levels and are set against published culture scores. The author's own assessment profile (DISC, Hogan Personality Inventory, Hogan Development Survey) is disclosed as part of the instrument, in the way a qualitative researcher discloses position.

1.5 Contribution and plan

The contribution is fivefold: the framework and its translation rules; the finch mapping and the breeder's-equation estimator; the empirical study of the measured style of embedded AI colleagues, which is to the author's knowledge the first to administer the instruments organisations use on people to AI colleagues configured as functions configure them; the nine-country lens with its two new observations (the advisor role tracks decision latency; predictability tracks codification); and a propositions table (Table 10.2) written so that it can later serve as the requirements list for a decision-support application, which is deferred to a second phase after the paper has been reviewed.

Chapter 2 reviews prior work and states the gap. Chapter 3 sets the rules for the evolutionary lens. Chapter 4 is the mapping chapter. Chapters 5 to 9 take the five levels in turn, with Chapter 7a reporting the empirical study inside the human-AI level. Chapter 10 integrates them, Chapter 11 draws implications for practice, Chapter 12 states limitations, and Chapter 13 concludes.

Table 10.1, in Chapter 10, sets out the ladder in one page: the five levels, what predicts behavior at each, the translation rule at each boundary, and the finch episode that illustrates it.

This chapter asked the question; Chapter 2 shows that five literatures each hold a piece of the answer and none holds the whole.

2. Prior work and the gap

2.1 Five literatures that do not cite each other

The question spans five bodies of work that have developed largely in isolation.

Multilevel organisational theory supplies the architecture. Kozlowski and Klein (2000) distinguish composition (a higher-level property that is the same kind of thing as its lower-level constituents, such as shared climate) from compilation (a higher-level property that is functionally equivalent but structurally different, such as team performance assembled from complementary contributions), and Chan (1998) gives five composition models (additive, direct consensus, referent-shift, dispersion, process) that specify how an individual-level construct becomes a group-level one. This literature tells us how to move between levels. It says little about what to measure at the bottom.

Team composition research supplies the bottom. Meta-analyses find that team-level means and minima of the Big Five predict team performance, with minimum agreeableness and mean conscientiousness the most consistent, and that the right aggregation rule depends on the trait (Barrick, Stewart, Neubert & Mount, 1998; Bell, 2007; Peeters et al., 2006). A more recent meta-analysis finds the effects small overall and heavily moderated (Han, Krieger, Kim, Nixon & Greiff, 2024). Collective intelligence research finds that interaction process explains about twice as much variance in group performance as member skill (Woolley, Chabris, Pentland, Hashmi & Malone, 2010; Riedl, Kim, Gupta, Malone & Woolley, 2021), with a well-known dispute over whether a single group factor exists (Credé & Howardson, 2017). This literature measures the bottom but stops at the team.

Human-AI teaming supplies the new member. The National Academies (2022) review calls explicitly for predictive performance models of human-AI teams; O'Neill, McNeese, Barron and Schelble (2022) find human-human teams almost always outperform human-autonomy teams in the experimental literature to date; Vaccaro et al. (2024) find, across 106 experiments, that human-AI combinations underperform the best single agent on average, with gains in creation tasks and losses in decision tasks. The trust literature (Lee & See, 2004; Dietvorst, Simmons & Massey, 2015; Logg, Minson & Moore, 2019; Bansal et al., 2021) explains why: reliance is miscalibrated in both directions. A parallel stream, starting in 2023, administers personality inventories to language models and finds measurable, conditionally stable, shapeable traits (Serapio-García et al., 2025; Bodroža, Dinić & Bojić, 2024), and the first experiments on human-AI personality pairing and on personality composition in all-AI teams appeared in 2025 and

2026 (Ju & Aral, 2025; Keluskar, Bhattacharjee & Liu, 2026; Kumar, Bharathwaj & Jurgens, 2026). This literature has a personality layer and a trust layer but no multilevel layer: nobody has yet placed the AI member inside a composition model.

Firm-level behavior supplies the fourth level. Upper-echelons theory (Hambrick & Mason, 1984) and its empirical descendants (Malmendier & Tate, 2008; Harrison, Thurgood, Boivie & Pfarrer, 2019) predict firm decisions from leader characteristics; text-derived culture predicts risk-taking and deal-making out of sample (Li, Mai, Shen & Yan, 2021); game-theoretic strategy (Brandenburger & Nalebuff, 1995) predicts behavior from the structure of the game rather than from any player's disposition; and the clockspeed and transformation literatures (Fine, 1998, 2026; van der Meulen, Weill & Woerner, 2020) predict a firm's posture from the speed of its environment and the order in which it handles organisational change. None of this connects to the personality of anyone below the top team.

Cross-national work supplies the fifth. Hofstede's dimensions (Hofstede, Hofstede & Minkov, 2010), GLOBE (House et al., 2004) and tightness (Gelfand et al., 2011) describe the environment; the liability of foreignness (Zaheer, 1995), institutional duality (Kostova & Roth, 2002) and CAGE distance (Ghemawat, 2001) describe what it costs to enter one. The effect sizes are modest (Taras et al., 2010) and the original data have been criticised (McSweeney, 2002). This literature moderates the levels below it but is rarely used to do so.

2.2 Practice has run ahead of theory

Commercial assessment vendors sell team-level products that aggregate individual profiles: team discovery maps, team reports, group culture reports. They implement a composition model, usually the additive one, without saying so, and none publishes outcome-predictive validation at the team level. The DISC family in particular, which is the practitioner entry point for much of the market, rests on an ipsative scoring format whose within-person logic does not support between-person prediction (Hicks, 1970; Meade, 2004), a limitation its own validity reports acknowledge (Assessments 24x7, 2024). The gap between what is sold and what is known is the practical motivation for this paper.

2.3 The gap

No existing work specifies the translation rule at each level boundary and follows it through from individual style to multinational posture. No existing work treats the AI member as a compositional element in the multilevel sense; Chapter 7a supplies the first measurement of that element with the instruments organisations already use. No existing work runs a single evolutionary case through all the levels to discipline the analogy. And no existing work tests a multilevel behavioral framework against a set of countries observed by one practitioner with one instrument. Table 2.1 summarises.

Table 2.1. Literature map

Level	What is measured	What is predicted	Best-evidence effect	Gap
Individual	Big Five, DISC, Hogan	Job performance;	r 0.2 to 0.3 for best trait	Ipsative tools sold as between-

Level	What is measured	What is predicted	Best-evidence effect	Gap
		form of behavior	(Sackett 2022)	person predictors
Human team	Aggregated traits; interaction process	Team performance	ρ 0.2 to 0.4 by trait and rule (Bell 2007); process about 2× skill (Riedl 2021)	Rule choice is theory, rarely stated
Human-AI team	Reliance, trust, task type; LLM "traits"	Combined performance	g about minus 0.2 on average (Vaccaro 2024)	No composition model includes the AI member (Chapter 7a measures it)
Firm	Leader traits, culture from text, game structure, clockspeed	Risk, deals, posture	Out-of-sample prediction from text (Li 2021)	Disconnected from levels below
Multinational	Culture dimensions; distance	Attitudes; entry cost	ρ about 0.18 (Taras 2010)	Used descriptively, not as moderator

The gap is established; Chapter 3 sets the rules for the lens the paper uses to close it.

3. The evolutionary lens: rules of use

3.1 Why biology

Management writing borrows from biology constantly and usually badly. Firms are said to evolve, industries to have ecosystems, ideas to go viral, and the borrowing rarely survives the question of what, precisely, is being claimed. This paper borrows heavily and therefore sets rules.

The reason to borrow at all is that biology has solved, formally, the problem this paper is about. The levels-of-selection question (does selection act on genes, individuals, groups, or species, and how do effects at one level translate to another) is the biological form of the multilevel problem in organisational theory, and it has a mathematics (the Price equation, the breeder's equation) and a body of field evidence that management lacks. Behavioral syndromes in animals (Sih, Bell & Johnson, 2004; Réale, Reader, Sol, McDougall & Dingemanse, 2007) are the biological form of

personality, with repeatability as the validation standard; the superorganism (Hölldobler & Wilson, 2009) and task allocation without central control (Gordon, 1996) are the biological form of team emergence; symbiosis is the biological form of a working relationship between unlike partners; population ecology (Hannan & Freeman, 1977) and business ecosystems (Moore, 1993; Iansiti & Levien, 2004) are the biological form of industry structure; and island biogeography is the biological form of the multinational's problem.

Charles H. Fine (1998), of MIT Sloan, gave this paper its licence. His argument for studying fruit flies was that the fastest-evolving species let the observer see a whole cycle in a human lifetime, so that the dynamics of slower industries could be read from the faster ones. The present paper extends the logic one step: if industries can be read from faster industries, organisations can be read from a population whose response to disturbance has been measured completely.

3.2 Three grades of analogy

Every biological analogy in this paper carries one of three grades, stated where it is used.

Shared mechanism: the same causal process operates in both domains. Heritability of personality (Bouchard & McGue, 2003) and the repeatability of animal temperament are the same phenomenon measured in different species. Reciprocal altruism and the tit-for-tat strategy (Axelrod & Hamilton, 1981) were published jointly by a political scientist and a biologist because the mechanism is the same. Analogies at this grade can be used to predict.

Structural isomorphism: different mechanisms, same structure. The finch's response to drought and an organisation's response to AI are driven by different processes (differential survival of genotypes; differential reward of routines), but both have the form variation, altered payoff, differential success, transmission, response. Analogies at this grade can be used to estimate, with the estimate labelled and its assumptions stated. The breeder's equation in Chapter 4 is used at this grade.

Metaphor: the analogy makes a point vivid but carries no inferential weight. The Big Bird hybrid lineage as a picture of a new human-AI working pattern is at this grade. Analogies at this grade are marked and are never used to support a proposition.

3.3 Where the analogy stops

Three differences between organisations and organisms bound every analogy in what follows. Organisations know they are being selected and change their behavior in response; finches do not. Organisations can change their traits within a generation; finches cannot. And organisations are selected partly by deliberate policy, including regulation, which has no biological equivalent. Chapter 4 returns to these limits after the case has been laid out, and Chapter 12 records them among the paper's limitations.

Table 3.1. The analogies used in this paper, by level and grade

Level	Business concept	Biological counterpart	Grade	Anchor sources
Individual	Behavioral style	Phenotype; behavioral	Shared mechanism	Sih et al. 2004; Réale et al. 2007;

Level	Business concept	Biological counterpart	Grade	Anchor sources
		syndrome		Bouchard & McGue 2003
Team	Emergent team behavior; early cooperation	Superorganism; task allocation; reciprocal altruism	Isomorphism (mechanism for cooperation)	Hölldobler & Wilson 2009; Gordon 1996; Axelrod & Hamilton 1981
Human-AI team	Reliance and advice-taking	Symbiosis spectrum	Metaphor	Licklider 1960
Firm	Routines, clockspeed, temporary advantage	Genes, generation time, Red Queen	Isomorphism	Nelson & Winter 1982; Fine 1998
Multinational	Ecosystems, entry into foreign markets	Population ecology; island biogeography; invasive species	Isomorphism	Hannan & Freeman 1977; Moore 1993; Zaheer 1995
All levels	Response to a disturbance	Selection, heritability, response ($R = h^2S$)	Isomorphism	Grant & Grant 1995, 2006

The lens is graded; Chapter 4 now runs one fully measured evolutionary case through a complete cycle of shocks.

Chapter 4. One organism, one cycle, one disturbing factor: the medium ground finch as a model of behavioral response to shock

4.1 Why a finch

The argument of this paper is that behavioral prediction degrades in a specific, describable way as one moves from the individual to the team, from the team to the firm, and from the firm to the multinational. To make that claim testable rather than merely plausible, the paper needs a reference case in which a population's response to a major disturbance has been observed at every level at once: in the individual (a measurable, heritable trait), in the population (survival, reproduction and the shift of the trait distribution), in the competitive system (the arrival of a rival that changes what the trait is worth), and in the wider geography (the islands from which immigrants come and to which they do not go). There is exactly one wild population for which all of this has been measured directly, over decades, by the same observers with the same instruments: the medium ground finch, *Geospiza fortis*, on the island of Daphne Major in the

Galápagos, studied by Peter and Rosemary Grant and their collaborators every year since 1973 (Grant & Grant, 2014).

The choice is not decorative. Chapter 3 committed the paper to grading each biological analogy as shared mechanism, structural isomorphism, or metaphor. The finch case is used here at the second grade. The mechanism by which beak depth responds to drought (differential survival of genotypes) is not the mechanism by which a project team responds to an AI copilot. But the structure of the response is the same: a population with heritable variation in a trait, a change in the environment that alters the payoff to that trait, differential success of variants, and a shift in the population that persists into the next generation. The breeder's equation that biologists use to predict the second step from the first, $R = h^2S$, has no biological content beyond that structure, and Section 4.6 applies it to organisations under exactly that reading.

Charles H. Fine of MIT Sloan, whose clockspeed framework anchors the firm-level chapters of this paper, opened his own book with fruit flies for the same reason: not because firms have chromosomes, but because "the fastest-evolving species allow us to see the whole cycle" (Fine, 1998). The Grants' finches are slower than *Drosophila* but faster than any organisation, and unlike fruit flies they were watched in the wild, under real weather, with real competitors. That is the reason they carry this chapter.

4.2 The cycle as observed

Baseline: heritable variation in a functional trait (1973–1976)

Daphne Major is a small volcanic island, roughly 0.34 square kilometres, with a resident breeding population of *G. fortis* in the low thousands and a second resident species, the cactus finch *G. scandens*. Beak depth in *G. fortis* is normally distributed, varies by roughly a third between the smallest and largest adults, and is strongly heritable: mid-parent–offspring estimates of h^2 for beak depth fall between 0.5 and 0.9 depending on the cohort and the correction for extra-pair paternity (Boag & Grant, 1978; Keller, Grant, Grant & Petren, 2001; Grant & Grant, 2002). Beak depth is functional: it sets which seeds a bird can crack in the time available, and the island's seed supply ranges from small soft seeds to the large, hard, spiny fruits of *Tribulus cistoides* (Grant & Grant, 2006).

In a normal year the trait is under weak or no selection. Birds with small beaks eat small seeds, birds with large beaks eat both, and the distribution stays where it is. The population is, in the language of Chapter 5, in a weak situation: individual variation is expressed, but it does not much matter for outcomes.

Negative shock: the 1977 drought

In 1977 the wet season failed. Twenty-four millimetres of rain fell where a normal season brings well over a hundred (Boag & Grant, 1981). The small soft seeds that most birds relied on were exhausted by mid-year and were not replenished. What remained were the large hard *Tribulus* fruits, and only birds with the deepest beaks could open them fast enough to survive.

The result was intense directional selection. The *G. fortis* population fell by about 85 per cent (Boag & Grant, 1981). The survivors were not a random sample: mean beak depth among the 85 survivors measured by Boag and Grant was 9.96 mm against 9.42 mm in the 642 birds measured

before the drought, a difference of roughly half a millimetre, or between four and six per cent depending on the cohort used as the baseline (Boag & Grant, 1981). Because the trait is heritable, the shift carried into the next generation: the offspring hatched in 1978 had beaks measurably deeper than the pre-drought generation, an evolutionary response observed within a single breeding cycle (Boag & Grant, 1981; Grant & Grant, 1995).

Three features of this episode matter for the organisational parallel. First, the environment did not select for "better" birds in any general sense; it selected for one trait that happened to match one resource. Second, the selection was invisible before the shock: nothing about the 1976 population predicted which birds would matter in 1977, except the trait itself. Third, the response was fast because the trait was heritable. Had beak depth been a purely developmental accident with no transmission to offspring, the survivors would have been larger-beaked but their children would not.

Positive shock: the 1982–83 El Niño

Five years later the environment reversed. The El Niño of 1982–83 brought 1,359 mm of rain to Daphne Major over nine months, ten times the previously recorded wet-season maximum (Gibbs & Grant, 1987b). Vegetation exploded, total seed biomass rose by an order of magnitude, and the composition of the seed supply flipped: small soft seeds went from about twenty per cent of seed biomass to more than eighty (Gibbs & Grant, 1987b). The large hard seeds that had been the only food in 1977 became a minor and inefficient option.

Selection reversed. In the drier years that followed, smaller birds with smaller beaks survived better, because they could process the now-abundant small seeds more efficiently than the large-beaked survivors of 1977 (Gibbs & Grant, 1987a). The trait that had saved the population in 1977 became a mild liability by 1985. The Grants titled the paper reporting this "Oscillating selection", and it is the empirical basis for a claim this paper will make repeatedly: the same trait is an asset or a liability depending on the regime, and a prediction of behavior that ignores the regime is a prediction about the wrong thing.

Competition shock: the 2004–05 drought with a new rival

The large ground finch, *G. magnirostris*, colonised Daphne Major in late 1982, carried in by the same El Niño. For twenty years it was a minor presence. Then came the droughts of 2003 (16 mm) and 2004 (25 mm), and this time the *Tribulus* seeds that had saved large-beaked *G. fortis* in 1977 were being eaten faster by the larger, more efficient *G. magnirostris* (Grant & Grant, 2006). The *G. fortis* population fell from 235 to 83 adults, the lowest since the study began. But the survivors this time were the smaller-beaked birds: beak size in the 2005 generation was 0.70 standard deviations below the 2004 generation, the strongest single-episode evolutionary change in the study's history (Grant & Grant, 2006). The Grants called this character displacement, the divergence of two competing species in the trait over which they compete, and it was the first time the process had been observed directly in the wild rather than inferred from its end state.

The lesson is different from 1977. In 1977 the environment alone set the payoff. In 2004 the environment created scarcity, but the payoff to the trait was set by the competitor. A large-beaked *G. fortis* in 2004 was competing not against the seed but against a bigger bird eating the

same seed, and lost. The trait that predicted survival was the one that avoided the rival, not the one that matched the resource.

Hybridisation: the Big Bird lineage

In 1981 an unusually large male finch arrived on Daphne Major. Genomic analysis decades later showed it to be a large cactus finch, *G. conirostris*, from Española, more than a hundred kilometres away (Lamichhaney et al., 2018). It bred with a resident *G. fortis* female. Its descendants were larger than either resident species, sang a song neither resident recognised, and from the third generation onward bred only among themselves. Within three generations, a reproductively isolated lineage occupied a niche on the island that neither parent species had filled (Lamichhaney et al., 2018). The Grants called them the Big Birds.

For the purposes of this paper the episode establishes that a new working form can arise from the combination of two existing ones, that it can do so within a handful of generations, and that it will be judged not by its parentage but by whether the niche it finds is viable. It also establishes that the combination was not designed; it was a contingent encounter that happened to work.

Mechanism found later

For thirty years the Grants measured beaks without knowing what made them. In 2015 and 2016 genome sequencing identified two loci: *ALX1*, associated with beak shape (blunt against pointed), and *HMGA2*, associated with beak size (Lamichhaney et al., 2015, 2016). The HMGA2 result closed the loop on the 2004–05 drought: the large-beak genotype at that locus carried a selection coefficient of 0.59 against it during the drought, meaning birds with that genotype were roughly forty per cent as likely to survive (Lamichhaney et al., 2016). The prediction the Grants had made from phenotype in 2006 was confirmed at the level of mechanism a decade later.

This last point matters for the epistemology of the whole paper. The Grants predicted evolutionary response from measured phenotype, heritability and selection differential, using the breeder's equation, without knowing the genes (Grant & Grant, 1995). Their predictions were good. Mechanism was not required for prediction; structure was. The paper argues the same for organisations: one does not need a theory of why a given DISC or Hogan profile behaves as it does to predict, within limits, what it will do, provided the trait is measured, the situation is characterised and the regime is known.

4.3 Artificial intelligence as the disturbing factor

The paper treats the arrival of AI colleagues in the workplace as a disturbance of the same structural kind as the events on Daphne Major: an environmental change that alters the payoff to traits people and organisations already have. This is a deliberate simplification. AI is not weather; it is adopted, governed, and to some degree designed. But from the standpoint of the people and teams inside an organisation in 2024 to 2026, the arrival of large language models in daily tools was experienced as exogenous, sudden and uneven, and the field evidence on its effects has the shape of a selection event.

The evidence base is reviewed in Chapter 7. What matters here is its shape. Generative AI raised the productivity of customer-support agents by about 14 per cent on average, with the largest gains going to the least experienced (Brynjolfsson, Li & Raymond, 2025). Among 758 consultants at a strategy firm, those working on tasks inside what the authors called the jagged technological frontier improved quality by around 40 per cent, while those working on a task outside the frontier and relying on the model were 19 percentage points less likely to produce a correct answer (Dell'Acqua et al., 2026). Across 106 experiments, human-AI combinations on average performed worse than the better of the human or the AI alone, with gains concentrated in creation tasks and losses in decision tasks (Vaccaro, Almaatouq & Malone, 2024). At the organisational level, a 2026 survey of over 1,000 executives found that 70 per cent of employees described themselves as ready to work with AI while only 27 per cent of leaders described their organisation as ready, and that organisational readiness explained roughly twice as much of the variance in value captured as personal readiness did (McKinsey, 2026).

Read as a selection event, this is a drought in which the seed supply changed rather than shrank. Some existing traits (novice status, willingness to iterate, a task portfolio inside the frontier) suddenly paid; others (expert status on a task the model handles badly, uncalibrated trust) suddenly cost. The distribution of who gained and who lost was set not by general ability but by the fit between the trait and the new regime, exactly as in 1977. And, as in 1983, the regime will move again: the frontier shifts with each model release, so a trait that pays in one release cycle may not pay in the next.

4.4 The mapping, level by level

Table 4.1 sets each finch episode against the level of the framework it best illustrates, the organisational parallel, and the evidence from the paper's sources. The table is the spine of the paper; the level chapters that follow (5 to 9) elaborate each row.

Table 4.1. Finch episodes mapped to framework levels

Finch episode	Biological structure	Framework level	AI-era parallel	Evidence
Baseline: heritable, functional beak variation under weak selection	Phenotype; heritability h^2 0.5–0.9; trait expressed but not consequential in a normal year	Individual (Ch. 5)	Stable behavioral style (DISC, Big Five, Hogan), partly heritable, measurable, consequential only when the situation is weak or the regime shifts	Bouchard & McGue 2003; Tett & Burnett 2003; Sackett et al. 2022; A247 validity report
1977 drought: 85% mortality, survivors 5% deeper-beaked,	Directional selection under scarcity; response	Individual and team (Ch. 5–6)	The jagged frontier: AI raises performance	Dell'Acqua et al. 2026; Brynjolfsson et al. 2025; Noy &

Finch episode	Biological structure	Framework level	AI-era parallel	Evidence
response inherited	proportional to $h^2 \times S$		inside the frontier and cuts it outside; who gains depends on the trait already held; novices gain most	Zhang 2023
1983 El Niño: abundance, selection reverses	Same trait, opposite payoff under a new regime	Team (Ch. 6)	Human-AI teams gain on creation tasks and lose on decision tasks; the same composition is an asset or a liability by task regime	Vaccaro et al. 2024; Riedl et al. 2021
2004–05 competitor arrival: character displacement	Payoff to the trait set by a rival, not by the resource	Firm (Ch. 8)	Incumbent facing AI-native rivals shifts posture; organisation clockspeed against industry clockspeed; the four explosions	Fine 2026; van der Meulen, Weill & Woerner 2020; Robertson 2026
Big Bird lineage: hybrid, isolated in three generations, new niche	Recombination of two existing forms into a viable third	Human-AI team (Ch. 7, 7a)	Centaur and cyborg working patterns; the reinvention horizon; hybrid intelligence	Dell'Acqua et al. 2026; McKinsey 2026
Archipelago: islands differ; immigration rare and consequential	Island biogeography; founder effects; each island a different selection regime	Multinational (Ch. 9)	Countries as islands; the foreign firm as immigrant; liability of foreignness; institutional duality; national culture as the island's seed	Zaheer 1995; Kostova & Roth 2002; Hofstede et al. 2010; Taras et al. 2010

Finch episode	Biological structure	Framework level	AI-era parallel	Evidence
			supply	
Genes found decades after the pattern	Prediction from phenotype and structure preceded mechanism	All levels (Ch. 10)	Behavior predicted from measured style and characterised situation without a theory of mechanism	Grant & Grant 1995; Fine 1998

Two rows deserve comment because they are where the analogy is most likely to be over-read.

The 1977 row equates the drought with the jagged frontier. The equation holds structurally: a change in the environment, a trait that suddenly matters, differential success. It does not hold in one important respect. The finches could not change their beaks. People can, to a degree, change their behavior, and organisations can change their processes. In Section 4.6 this appears as the difference between h^2 in a finch (a fixed fact of inheritance) and the "heritability" of organisational routines, which is partly a design choice. The analogy therefore sets an upper bound on how fast and how far an organisation can be expected to move, not a prediction of where it will end up.

The Big Bird row equates hybrid speciation with the emergence of stable human-AI working patterns. Here the analogy is weaker still, and the paper marks it at the third grade, metaphor. What the Big Birds show is that a viable new form can arise quickly from recombination and be judged by its niche rather than its origin. They do not show that such forms are designed, governed, or replicable, all of which are true of organisational hybrids and false of finches.

4.5 The figures

Figures 4.1a and 4.1b draw the Daphne Major record from 1973 to 2006 on one timeline: wet-season rainfall for the four years in which the sources report it, the adult population of *G. fortis* and *G. magnirostris* for 1997 to 2006 and the 1976 to 1978 collapse, and mean beak size (the first principal component of beak dimensions, sexes combined) for every year from 1973 to 2006, with the four episodes marked. Both panels are drawn schematically from Figures 2 and 3 of Grant and Grant (2006), anchored to the values the sources print: the earlier population figures and the 1977 rainfall are from Boag and Grant (1981), the 1983 rainfall from Gibbs and Grant (1987b), and the years between the printed anchors follow the published curves' shape rather than digitised values. The Grants did not publish counts for 1979 to 1996 in that paper because *G. magnirostris* was then too scarce to compare, and the panel leaves those years blank rather than interpolate. The standardised selection differentials printed on the figure are the Grants' own: 0.64 (males) and 0.67 (females) in 1977, and, in the opposite direction, 0.77 and 0.65 in 2004 (Grant & Grant, 2006, Table 2). The figure makes visible what the text can only assert: the 1978 jump, the slow twenty-year drift back toward the 1973 mean after the El Niño, the stability of the 1990s, and the single-year drop in 2005 that was larger than anything before it.

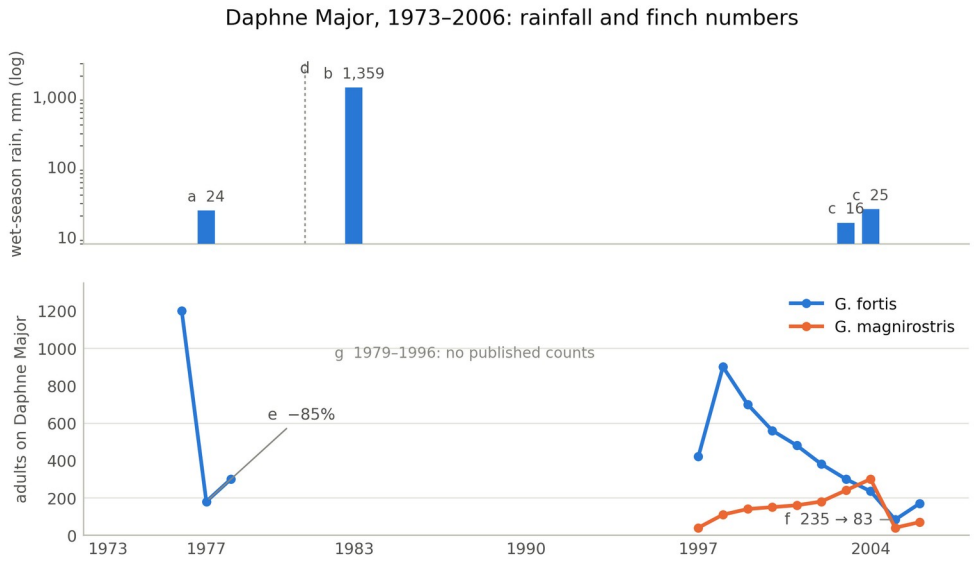


Figure 4.1a. Daphne Major, 1973 to 2006: wet-season rainfall and adult finch numbers, drawn schematically from the sources' published anchors. a: 1977 drought, 24 mm (Boag & Grant 1981). b: 1982-83 El Niño, 1,359 mm, ten times the previous maximum (Gibbs & Grant 1987b). c: 2003 and 2004, 16 and 25 mm (Grant & Grant 2006). d: 1981, arrival of the immigrant G. conirostris male that founded the Big Bird lineage (Lamichhaney et al. 2018). e: about 85 per cent mortality of G. fortis in 1977 (Boag & Grant 1981). f: 235 to 83 adults in 2005, the lowest since 1973 (Grant & Grant 2006). g: 1979 to 1996 not shown; the source publishes no counts before 1997.

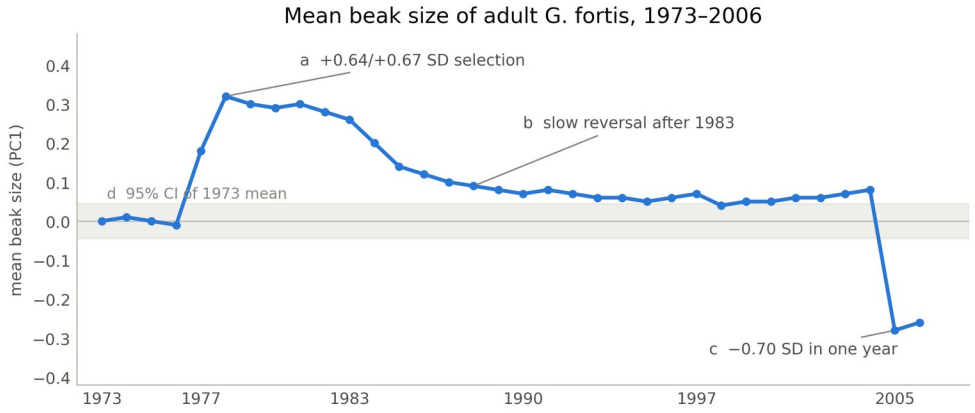


Figure 4.1b. Mean beak size of adult G. fortis (first principal component, sexes combined), 1973 to 2006, drawn schematically after Grant & Grant 2006, Fig. 2, anchored to the marked values. a: 1978 survivors and their offspring larger; standardised selection differential +0.64 (males) and +0.67 (females). b: slow reversal toward smaller beaks after the 1983 El Niño (Gibbs & Grant 1987a). c: 2005 generation 0.70 SD smaller; selection differential minus 0.77 (males) and minus 0.65 (females); the strongest change in the study (Grant & Grant 2006). d: shaded band marks the 95 per cent confidence limits of the 1973 mean.

Figure 4.2 draws the AI shock on the same layout for the period 2022 to 2026. The resource axis is a capability proxy (the release cadence of frontier models). The population axis is the distribution of organisations across the three horizons reported by McKinsey (2026): adoption,

impact and reinvention, with reinvention at 11 per cent. The trait axis is the estimated share of knowledge-work tasks inside the frontier, taken from the task-level results in Dell'Acqua et al. (2026) and Noy and Zhang (2023). The figure is labelled as an estimate and its assumptions are printed on it and in Appendix B.

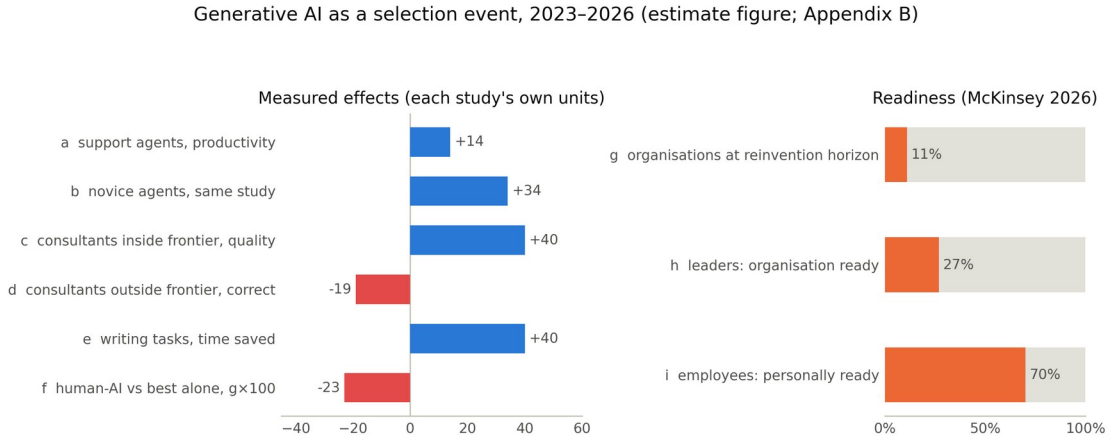


Figure 4.2. Generative AI as a selection event, 2023 to 2026. Panel 1, measured effects, in each study's own units, so sign and rough size are the point: a, customer-support agents, +14 per cent productivity (Brynjolfsson, Li & Raymond 2025); b, novice agents in the same study, +34 per cent; c, consultants on a task inside the model's frontier, +40 per cent quality (Dell'Acqua et al. 2026); d, consultants on a task outside the frontier, minus 19 percentage points correct; e, writing tasks, about 40 per cent faster (Noy & Zhang 2023); f, human-AI combinations against the best single agent across 106 experiments, Hedges' g about minus 0.23 (Vaccaro, Almaatouq & Malone 2024). Panel 2 (McKinsey 2026): g, organisations at the reinvention horizon, 11 per cent; h, leaders saying their organisation is ready for AI, 27 per cent; i, employees saying they are ready, 70 per cent. Estimate figure; see Appendix B.

Figures 4.3a and 4.3b are the selection-and-response diagram, drawn twice: once for the finches with measured values of S and h^2, once for an organisation with S and h^2 defined as in Section 4.6.

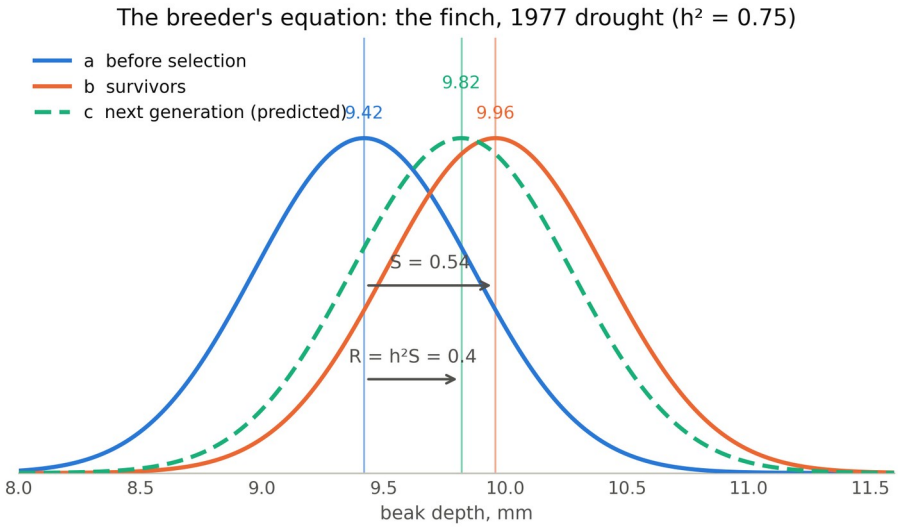


Figure 4.3a. The breeder's equation for the finch, 1977 drought. a: population before selection, mean beak depth 9.42 mm. b: survivors, mean 9.96 mm; S is the selection differential, 0.54 mm. c: next generation, predicted mean 9.82 mm; R = h²S with h² = 0.75, giving 0.41 mm. Sources: Boag & Grant 1981; Keller et al. 2001; Grant & Grant 1995.

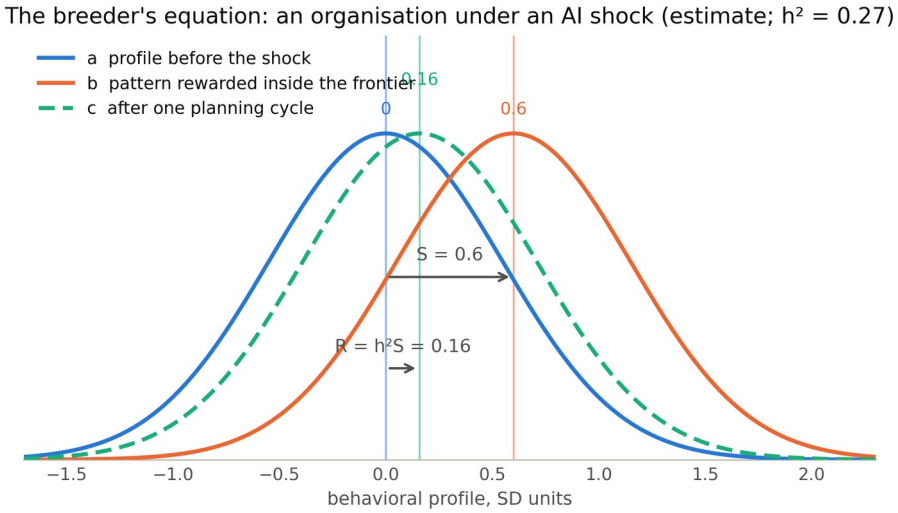


Figure 4.3b. The breeder's equation for an organisation under an AI shock, an estimate. a: behavioral profile before the shock. b: the pattern rewarded inside the model's frontier; S about 0.6 SD, from the inside-versus-outside-frontier gap in Dell'Acqua et al. 2026. c: profile after one planning cycle; R = h²S with h² = 0.27, the share of leaders describing their organisation as ready (McKinsey 2026), giving 0.16 SD. Assumptions in Appendix B.

Figure 4.4 shows character displacement as two beak-size distributions before and after the arrival of *G. magnirostris*, and beneath it the same shape for an incumbent and an AI-native rival on a decision-latency axis, using Fine's (2026) estimate that leading Chinese automakers bring a vehicle to market in roughly twenty months against about forty for established firms.

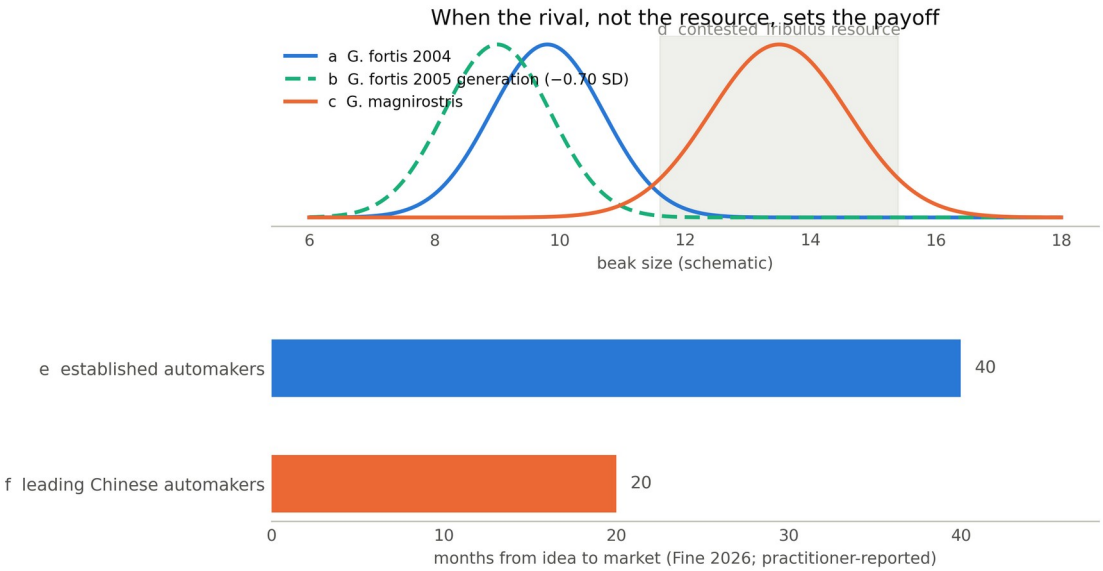


Figure 4.4. When the rival, not the resource, sets the payoff to a trait. Panel 1, schematic from Grant & Grant 2006: a, G. fortis in 2004; b, the 2005 generation, 0.70 SD smaller; c, G. magnirostris, larger-beaked and faster at eating Tribulus seeds; d, the contested Tribulus resource. Panel 2, from Fine 2026: e, established automakers, about 40 months from idea to market; f, leading Chinese automakers, about 20 months. Practitioner-reported figures, not a statistical test.

Figure 4.5 is the archipelago: the nine countries of Chapter 9 positioned by power distance and uncertainty avoidance, sized by the weight the author assigns them, with arrows for the author's own movements between them. It is introduced here and used again in Chapter 9.

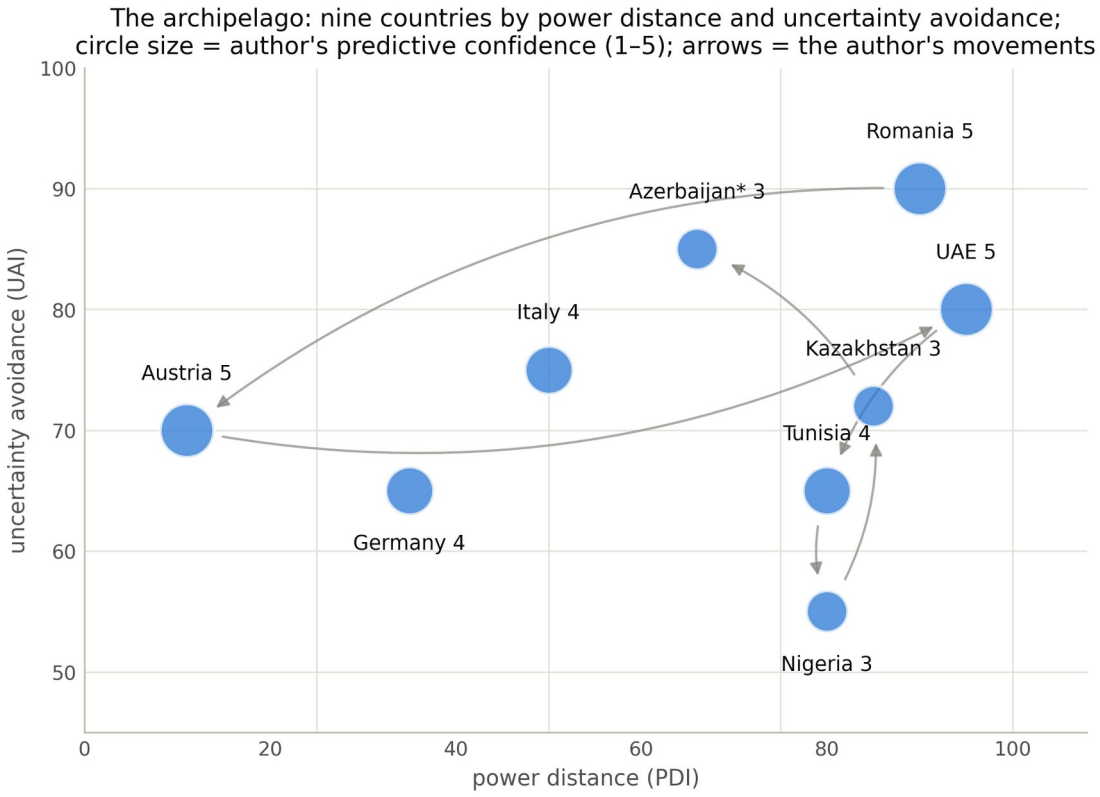


*Figure 4.5. The archipelago: the nine countries of Chapter 9 by Hofstede power distance and uncertainty avoidance, circle size showing the author's confidence in predicting workplace behavior there (1 to 5, the number after each name). Scores from The Culture Factor, retrieved August 2026; Austria, Italy, Germany and the Arab cluster come from the original IBM data, the rest are later estimates. *Azerbaijan has no published score and is placed at Turkey's values on the author's judgement.*

4.6 An estimation tool: the breeder's equation for organisations

Biologists predict the response of a population to one generation of selection with the breeder's equation:

$R = h^2 S$

where S is the selection differential (the difference between the mean trait of the individuals who survive and reproduce and the mean of the population before selection), h^2 is the narrow-sense heritability of the trait (the proportion of trait variance transmitted from parents to offspring), and R is the response (the difference between the mean of the next generation and the mean of the previous one). Grant and Grant (1995) used it on Daphne Major and found that it predicted the observed responses well.

The paper proposes the same equation as a first-order estimator of how much an organisation's behavioral profile will move in one planning cycle in response to a disturbance such as AI adoption. The terms are read as follows.

S, the selection differential, is how much the disturbance discriminates between behavioral variants. A regime in which AI-augmented work is rewarded and the rest is not has a large S; a regime in which everyone is paid the same regardless has S near zero. Fine's (2026) industry clockspeed is the natural proxy: fast-clockspeed industries discriminate hard and often, slow ones rarely. Where a direct measure is wanted, the difference in outcomes between AI-inside-the-frontier and AI-outside-the-frontier workers in Dell'Acqua et al. (2026) gives an S of about 0.6 standard deviations for one task regime, close to the standardised selection differentials the Grants measured in both finch episodes (0.64 to 0.67 in 1977; 0.65 to 0.77 in 2004; Grant & Grant, 2006, Table 2).

h^2, organisational heritability, is the proportion of a behavioral pattern that is transmitted from one cycle to the next through routines, training, decision rights and culture, rather than reset each time. Nelson and Winter (1982) called routines the genes of the firm; the term is used here in that sense. An organisation that reorganises every five months, as in the Romanian refinery case in Chapter 8, has low h^2: whatever the last cycle selected for is lost in the next restructuring. An organisation with stable decision rights and documented practice has high h^2. The McKinsey (2026) organisational-readiness figure of 27 per cent is offered as a rough population-level proxy for h^2 in 2026, with the caveat that it measures self-reported readiness, not transmission.

R, the response, is the predicted shift in the organisation's behavioral profile after one cycle.

The equation is useful less for its numbers than for what it separates. It says that a strong selection pressure (an aggressive AI mandate) will produce no lasting change in an organisation with low heritability, because nothing carries forward; and that a highly heritable organisation under weak pressure will not move either, because nothing selects. It also says that response is bounded by the product, so the fastest possible organisational change under a given disturbance is set jointly by how hard the environment pushes and how well the organisation retains what the push selected. These are the two levers a leader controls, and they correspond to Fine's (2026) two recommendations: calibrate speed to the environment, and treat speed as a governed capability.

Appendix B works the equation for three illustrative organisations drawn from the country data of Chapter 9 (high S, low h^2: the Romanian refinery; low S, high h^2: the Austrian headquarters; high S, high h^2: the Gulf state champion) and shows that the predicted ordering of response matches the author's field observation of which organisations actually changed their working patterns fastest. This is a consistency check, not a test; the numbers are estimates and are labelled as such.

4.7 What the finches do not tell us

The chapter closes with the limits of its own case. Finches do not choose their traits, plan their responses, or read papers about themselves. Organisations do all three, and the second-order effect of a population that knows it is being selected has no analogue on Daphne Major. Finches have one heritable trait under selection at a time, or at most a few; organisations have many, entangled. Selection on Daphne Major was applied by weather and by a competitor that arrived by chance; selection on organisations is applied partly by markets and partly by deliberate policy, including the regulation reviewed in Chapter 7. And finches cannot leave the island by

choice; the Tunisian case in Chapter 9, in which a foreign operator's response to a stalled environment was to prepare an exit, has no finch equivalent.

These limits are the reason the paper grades the analogy at structural isomorphism and not at shared mechanism. What survives the limits is the structure: variation, differential payoff under a changed regime, transmission, response, and a competitor that resets the payoff. That structure is enough to organise the five level chapters that follow, and enough to make the estimation in Section 4.6 more than a metaphor.

The finch has supplied the structure and the estimator; Chapters 5 to 9 climb the five levels with both in hand.

Chapter 5. Level 1: the individual

5.1 What the instruments measure

Three families of instrument dominate the assessment of individual workplace behavior, and they do not measure the same thing.

The first is the Big Five, the factor-analytic model of personality that underlies most scientific work and the Hogan inventories used in this paper. It is normative: a person's score on conscientiousness is a position relative to a reference population, and two people's scores can be compared. Its predictive validity for job performance is the best-documented finding in personnel psychology and also the most revised. Schmidt and Hunter's (1998) synthesis put conscientiousness at 0.31; Sackett, Zhang, Berry and Lievens (2022), correcting for what they showed to be over-corrected range restriction in the earlier work, put it at 0.19 to 0.20, with structured interviews at 0.42 and general mental ability at 0.31. The numbers are small in absolute terms. A correlation of 0.2 is four per cent of variance. They are also stable, replicated, and larger than most alternatives that can be measured before hiring.

The second family is DISC. It descends from Marston's (1928) theory of the emotions of normal people, which proposed four modes of response (dominance, inducement, submission, compliance) to a favourable or antagonistic environment; Marston built no instrument, and the questionnaires that carry his labels were developed by others from the 1950s onward. Modern DISC instruments are typically ipsative: the respondent chooses, in each block, which of four statements is most and least like them, so that the four scale scores are forced to trade off against one another and sum to a constant. This format has a known consequence. Ipsative scores describe the ordering of traits within a person and cannot support comparisons between people (Hicks, 1970; Meade, 2004). They are not a defect in the instrument; they are a different kind of measurement, suited to self-insight and coaching and unsuited to selection or between-person prediction. The validity report for the instrument used in this paper acknowledges the point in its own introduction, warning that normative statistical techniques should be applied with caution to ipsative data (Assessments 24x7, 2024). The same report shows the trade-offs: in a sample of ten thousand, dominance and steadiness correlate at minus 0.55 and influence and conscientiousness at minus 0.57. Independent evaluations find that DISC dimensions are better described as combinations of Big Five factors than as four constructs (Martinussen, Richardsen & Vårum, 2001) and that the evidence base for behavioral-style tools rests more on

perception than on documented outcomes (McKenna, Shelton & Darling, 2002). Where forced-choice formats have been made to yield normative information, through partially ipsative designs or Thurstonian item-response models, criterion validity for conscientiousness rises to around 0.4 (Salgado, Anderson & Táuriz, 2015; Brown & Maydeu-Olivares, 2013). Fully ipsative four-scale profiles are the weakest configuration.

The third family is the derailer inventory, of which the Hogan Development Survey is the standard example. It measures not typical behavior but behavior under stress, fatigue or reduced self-monitoring, on eleven scales that map loosely onto the personality disorders and, in a work context, onto the ways a strength becomes a liability. It is normative and it measures a regime the other two instruments do not: the same person in a drought.

5.2 When traits predict

Whether a trait is expressed depends on the situation. Trait activation theory (Tett & Burnett, 2003) holds that a trait predicts behavior when the situation contains cues relevant to it and rewards its expression. Situational strength (Meyer, Dalal & Hermida, 2010) is the general form: a strong situation, defined by clarity of expectations, consistency of cues, constraints on behavior and consequences for deviation, suppresses individual differences; a weak one releases them. Judge and Zapata (2015) showed that Big Five validity for job performance is substantially higher in weak situations, with the effect concentrated in jobs that allow discretion.

This is the individual-level form of the finch's baseline. In a normal year, beak depth varies and nothing depends on it; in a drought, it decides survival. The trait is stable across both; the situation sets its consequence. The organisational corollary is that a regulated pipeline control room is a strong situation in which the operator's style will barely show, and a first negotiation with a foreign state partner is a weak one in which it will dominate.

A second boundary is what the trait predicts. The DISC in Action material used in practitioner training (Assessments 24x7, 2025) describes, for each style, how the person approaches a problem, what they prioritise and how they should be approached in return: the D style is direct and results-first and wants the purpose stated up front; the C style asks specific questions after observing and wants the process detailed; the S style seeks consensus and wants time to prepare; the I style wants recognition and room to brainstorm. These are predictions about the form of behavior, and they are the kind of prediction the instrument is good at. They say nothing about whether the problem gets solved. The distinction runs through the paper: style predicts form reliably and outcome weakly, and the ratio between the two narrows as the situation strengthens.

5.3 The author's profile as a worked case

The paper's field lens (Chapter 9) rests on one observer, and the observer's profile is therefore disclosed as part of the instrument. The author holds three assessments: a DISC leadership report (Assessments 24x7, 2026), and the Hogan Personality Inventory and Hogan Development Survey, both administered as part of a professional leadership assessment in February 2025.

The DISC report gives two profiles. The adapted style, which the instrument describes as the respondent's perception of the behavior the current role requires, is Id: dominance 59,

influence 80, steadiness 41, conscientiousness 41. The natural style, described as instinctive behavior that surfaces under stress and is stable across roles, is Ic: dominance 48, influence 70, steadiness 34, conscientiousness 52. The instrument labels the combined pattern “Assessor” and characterises it as high orientation toward both people and quality control.

The Hogan Personality Inventory places the author, against a leader norm group, at the 97th percentile on ambition, the 90th on adjustment, the 88th on learning approach, the 72nd on interpersonal sensitivity, the 67th on sociability, the 58th on prudence and the 55th on inquisitiveness. The Hogan Development Survey puts cautious and bold both at the 75th percentile (moderate risk), reserved at 69 and colourful at 66 (low risk), and all other scales at low or no risk, with dutiful the lowest at 26. The scale definitions and the interpretive text are the publisher’s (Hogan Assessment Systems, 2025) and are not reproduced here; the percentiles are sufficient for the argument.

Three observations follow that the paper uses later.

First, the two instruments converge from different directions. The DISC adapted profile lifts dominance by eleven points and drops conscientiousness by eleven relative to the natural profile; the Hogan places ambition in the top three per cent. An instrument that measures within-person ordering and an instrument that measures between-person position both describe a person whose role pulls toward drive and away from procedure. That is a small piece of convergent validity, and it is also a caution: the DISC gap between adapted and natural is the instrument’s own record that the situation is pulling on the trait.

Second, the Hogan profile contains a tension the DISC cannot see. Cautious and bold are opposite derailers, one an excess of restraint and the other an excess of confidence, and both are elevated. A single-score summary would average them to nothing. The derailer inventory shows that the same person has two distinct stress modes and that which one appears depends on the situation. Chapter 9 reports that the nine countries triggered them differently: the bold side in Romania, the cautious side in Nigeria and Tunisia.

Third, none of this predicts an outcome. The profile says what form the author’s behavior takes, in which direction the current role bends it, and how it fails under load. It does not say whether a given project was delivered, and the paper does not claim it does.

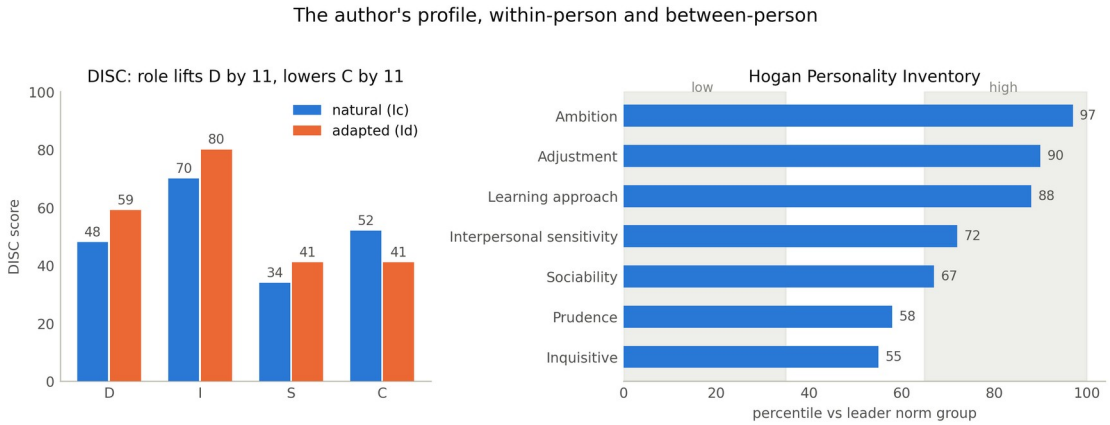


Figure 5.1a. The author's profile, within-person and between-person. Panel 1, DISC natural (Ic: D48 I70 S34 C52) and adapted (Id: D59 I80 S41 C41) styles; the role lifts D by 11 and lowers C by 11 (Assessments 24x7, 2026). Panel 2, Hogan Personality Inventory percentiles against a leader norm group, shaded bands low (under 35) and high (over 65) (Hogan Assessment Systems, 2025).

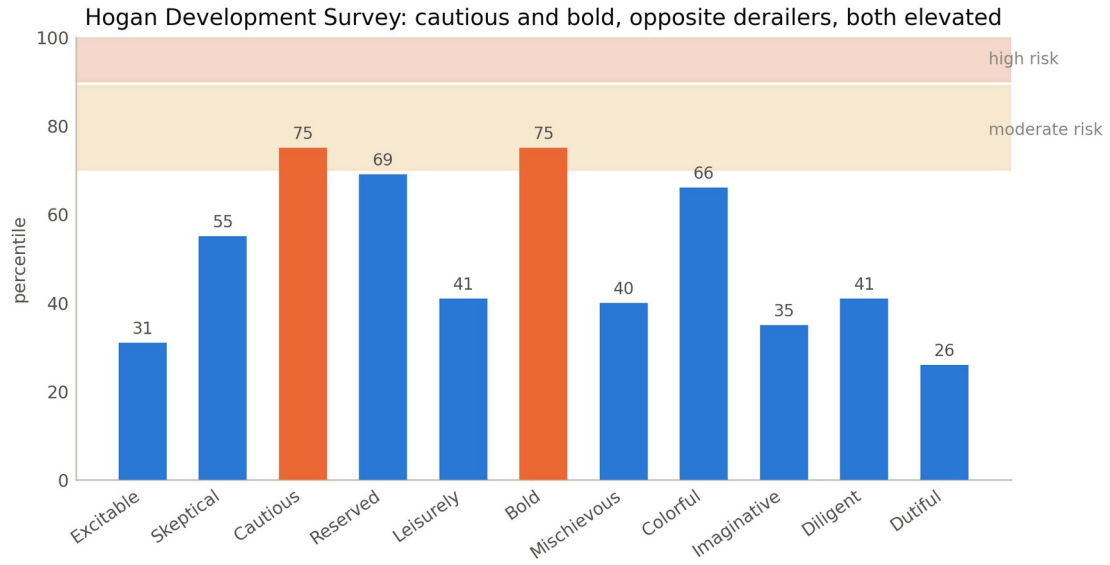


Figure 5.1b. The author's Hogan Development Survey: behavior under stress, percentiles with the publisher's risk bands. Cautious and bold, opposite derailers, are both at the 75th percentile (moderate risk); all other scales are at low or no risk. Disclosed as instrument calibration for the field lens of Chapter 9; the profile predicts the form of behavior and its stress modes, not outcomes.

5.4 Biology at this level

The analogy at the individual level is at the first grade, shared mechanism. Personality in humans is partly heritable, at roughly 40 to 50 per cent for the Big Five (Bouchard & McGue, 2003), and its expression is environment-dependent. Behavioral ecologists have found the same structure in animals: consistent individual differences in boldness, exploration, activity, aggressiveness and sociability that persist across contexts and time (Sih, Bell & Johnson, 2004;

Réale, Reader, Sol, McDougall & Dingemanse, 2007). The standard of validation in that literature is repeatability, the share of variance in a behavior that is between individuals rather than within them, and it is the direct analogue of test-retest reliability. The finch's beak is a morphological trait rather than a behavioral one, but the point transfers: a heritable, stable, measurable characteristic that is neutral in one regime and decisive in another.

5.5 Propositions

P1. Individual style predicts the form of workplace behavior reliably and its outcome weakly; the ratio between the two shrinks as situational strength rises.

P2. Ipsative profiles such as DISC describe within-person tendencies and support prediction of form for that person; between-person forecasts require normative or item-response-scored instruments, and a derailer inventory is required to predict behavior under stress at all.

The individual predicts form, not outcome; Chapter 6 asks what survives when individuals become a team.

Chapter 6. Level 2: the human team

6.1 The translation problem: emergence

A team has no personality. It has members with personalities, and whatever is said about the team's behavior is a statement about how those members' characteristics combine. Multilevel theory calls this emergence and distinguishes two ideal types (Kozlowski & Klein, 2000). Under composition, the team-level property is the same kind of thing as the individual-level one, aggregated: a team's mean conscientiousness is conscientiousness, summed. Under compilation, the team-level property is functionally equivalent but structurally different: team performance is assembled from complementary contributions that are not the same as one another and cannot be averaged. Chan (1998) gives five composition models that specify the rule. The additive model takes the mean regardless of agreement; direct consensus takes the mean only if members agree; referent-shift consensus rewrites the items so that the team, not the individual, is the referent; the dispersion model treats the variance itself as the construct; and the process model treats the team-level property as an analogue of an individual process rather than an aggregate.

The choice of rule is a theoretical claim, and the empirical literature shows that it decides whether personality predicts anything at the team level.

6.2 What composition predicts

Bell's (2007) meta-analysis of 225 correlations found that in laboratory teams personality composition predicted almost nothing, and in field teams it predicted modestly and only when the aggregation matched the trait: the team's mean conscientiousness, the team's minimum agreeableness, and the mean of collectivism and preference for teamwork were the meaningful predictors, at corrected correlations in the range of 0.2 to 0.4. Barrick, Stewart, Neubert and Mount (1998), in 51 manufacturing teams, had already shown the minimum rule at work: one

member very low on agreeableness, or high variance in conscientiousness, was enough to depress performance, the origin of the one-bad-apple argument. Peeters, van Tuijl, Rutte and Reymen (2006) found that elevation in agreeableness (0.24) and conscientiousness (0.20) helped and that variability in conscientiousness (minus 0.24) hurt. Han, Krieger, Kim, Nixon and Greiff (2024), updating the field, found the effects small overall and heavily moderated by team type and task.

Two things follow. First, the mean is the wrong rule for cooperation-relevant traits: a team's capacity to cooperate is set by its least cooperative member, not its average. Second, even with the right rule the effect is small. Personality composition is a real but minor input.

6.3 What interaction predicts

The larger input is what happens once the team starts working. Woolley, Chabris, Pentland, Hashmi and Malone (2010) found a general collective-intelligence factor that explained 30 to 50 per cent of variance in group performance across diverse tasks, weakly related to the average or maximum intelligence of members and strongly related to the average social sensitivity of members, the equality of conversational turn-taking, and the proportion of women in the group. Credé and Howardson (2017) disputed the single-factor model on re-analysis; Riedl, Kim, Gupta, Malone and Woolley (2021), pooling 22 studies and 1,356 groups, found the factor robust and estimated that group collaboration process was about twice as important as individual member skill in predicting it. Edmondson (1999), in 51 teams at one manufacturer, found team psychological safety, a shared belief that the team is safe for interpersonal risk-taking, correlated at 0.72 with performance, with learning behavior fully mediating the link.

Interaction process is itself partly path-dependent, and the path begins early. Axelrod and Hamilton (1981) showed that cooperation is stable in a repeated game when the probability of future interaction is high enough, that the strategy which won both of Axelrod's tournaments was the simplest one (cooperate first, then reciprocate), and that its winning properties were that it was never the first to defect, retaliated promptly, forgave quickly and was legible to the other side (Axelrod, 1984). Fischbacher, Gächter and Fehr (2001) found that about half of experimental subjects are conditional cooperators who match what others do, and that contributions in public-goods games decay as conditional cooperators revise their expectations downward. The negotiation-teaching exercise cited in Chapter 1 shows the same mechanism at classroom scale: round-one cooperation predicts early relationship quality, which predicts later cooperation, which predicts the bonus rounds and the final score (Emeritus, 2026). A team's trajectory is set disproportionately by its first moves.

6.4 The advisor inside the team

One member of many teams is not a peer but an advisor, and Salacuse (2016) supplies the only structured account of that role. Advice is a communication intended to help another person choose a course of action; it is not binding, but information asymmetry produces dependence over time. Salacuse identifies three models of the advisor-client relationship: the director, who takes control and steers the client; the servant, who responds to the client's demands and holds no initiative; and the partner, who manages the problem jointly with the client while the client keeps final authority. In his survey of 71 advisors at the European Union Council, 80 per cent preferred the partner model and fewer than 3 per cent the director model; the relationship, not

the information, was what made advice effective, and advising style varied by region in ways that recur in Chapter 9 (Eastern European advisors leaned toward delivering information, higher time sensitivity and lower emotional expression).

The paper treats the advisor's assigned role as a composition variable in its own right. A team that casts its advisor as a servant is a different team from one that casts the same person as a partner, and the difference shows in decision latency and in who is allowed to say no. Chapter 7 extends this to the AI member, and Chapter 9 reports that the role assigned to the author across nine countries tracked the organisation's decision system rather than its national culture.

6.5 Biology at this level

The analogy is structural isomorphism, with one component (reciprocal cooperation) at the grade of shared mechanism. Social insect colonies behave predictably as wholes although no individual carries a plan: task allocation in ant colonies adjusts to conditions through local interactions and simple rules, with no central control (Gordon, 1996), and the colony as a unit of selection is the subject of Hölldobler and Wilson's (2009) superorganism. Couzin, Krause, Franks and Levin (2005) showed that a small informed minority can steer a moving group's decision and that the required proportion shrinks as the group grows. These are the biological forms of two findings above: that process predicts more than composition, and that one member can set the group's direction. Axelrod and Hamilton's paper was co-authored by a biologist because the tit-for-tat result applies without change to reciprocal altruism in animals; that component of the analogy is not an analogy.

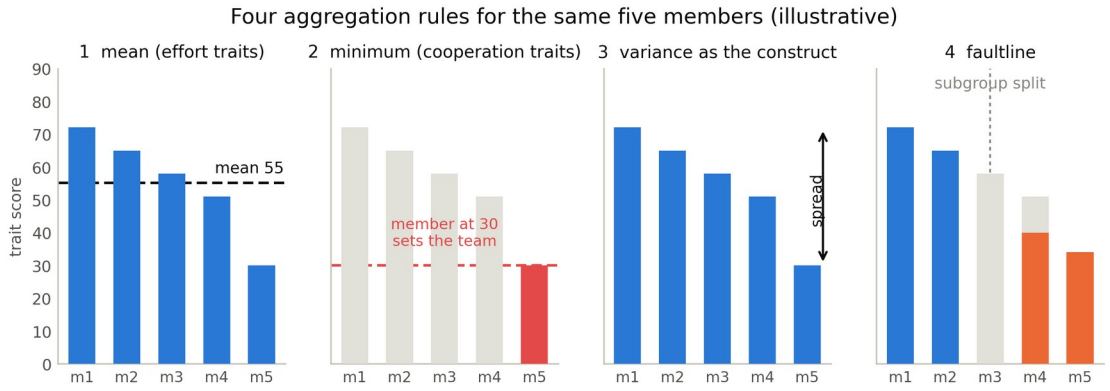


Figure 6.1. Four aggregation rules for the same five members (illustrative scores). 1, the mean, which fits effort traits such as conscientiousness; 2, the minimum, which fits cooperation traits such as agreeableness, where one member at 30 sets the team's capacity; 3, the variance, where heterogeneity itself is the risk; 4, a faultline, where the team splits into subgroups. After Chan 1998; Bell 2007; Barrick et al. 1998; Peeters et al. 2006.

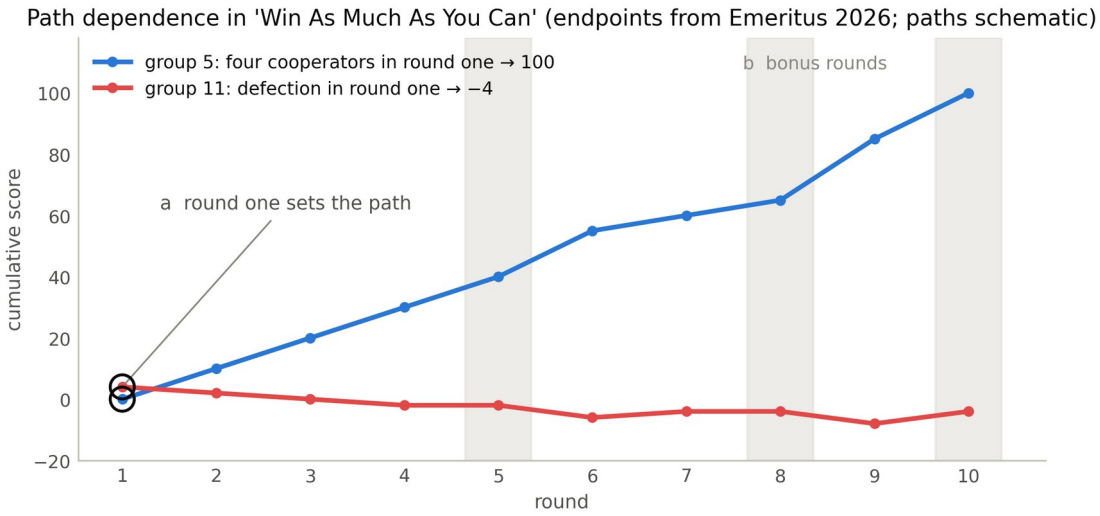


Figure 6.2. Path dependence in the 'Win As Much As You Can' exercise. a: round one, the choice that sets the path. b: bonus rounds 5, 8 and 10, shaded. Final scores (100 for group 5, four cooperators in round one; minus 4 for group 11, defection in round one) are from the cohort data in Emeritus 2026; the intermediate paths are schematic, drawn to the exercise's payoff rules.

6.6 Propositions

P3. Team behavior is predicted better by the aggregation rule matched to the trait than by any single summary statistic; minimum-based rules dominate for cooperation-relevant traits, mean-based rules for effort-relevant traits, and dispersion rules where heterogeneity itself is the risk.

P4. Early interaction behavior predicts later team behavior more strongly than pre-team trait composition does; the first rounds are the highest-leverage point of intervention.

P5. Psychological safety and clarity of the advisor's assigned role are emergent states that mediate the link between composition and behavior; where they are absent, composition predicts little.

Teams translate traits through aggregation rules and first moves; Chapter 7 adds the member that is not a person.

Chapter 7. Level 3: the human-AI team

7.1 What changes when one member is a machine

A team that includes an AI system as a working member differs from a human team in three ways that matter for prediction. The AI member gives advice rather than taking decisions, so

the relationship between it and the humans is an advisory one in Salacuse's (2016) sense: its output is not binding, the information asymmetry is large and grows, and dependence follows. The AI member's behavioral style, to the extent it has one, is not a stable property of a person but a function of the model version, the system prompt and the task. And the humans' reliance on it is the main channel through which its presence changes the team's behavior. Each of these has a literature, and the three together form the level.

Seeber et al. (2020) set the definitional bar: a machine is a teammate when it holds a role, shares goals and interacts with the humans as a participant rather than a tool. The National Academies (2022) review of human-AI teaming closes by calling for predictive models of team performance that include the AI member, a call that had not been answered in the multilevel sense at the time of writing. The intended outcome has a name in that literature, hybrid intelligence: combined performance that neither party could reach alone.

7.2 The net effect is contingent

The experimental record does not support the assumption that adding an AI member improves a team. O'Neill, McNeese, Barron and Schelble (2022), reviewing 76 studies, found that human-human teams almost always outperformed human-autonomy teams. Vaccaro, Almaatouq and Malone (2024), meta-analysing 106 experiments, found that human-AI combinations on average performed worse than the better of the human or the AI alone, with an average effect around minus 0.2 in standardised units; gains appeared in creation tasks, losses in decision tasks, and the combinations that did outperform were those in which the human alone had been better than the AI alone. The field evidence points the same way with an added twist. Brynjolfsson, Li and Raymond (2025) found a 14 per cent average productivity gain among customer-support agents given a generative assistant, rising to 34 per cent for the least experienced. Dell'Acqua et al. (2026), with 758 consultants, found quality gains of around 40 per cent on tasks inside the model's frontier and a 19 percentage-point fall in correctness on a task outside it, and identified two working patterns among the successful: centaurs, who divide tasks between themselves and the model, and cyborgs, who interleave. Noy and Zhang (2023) found writing tasks completed about 40 per cent faster with a quality gain, the gain again concentrated among weaker writers.

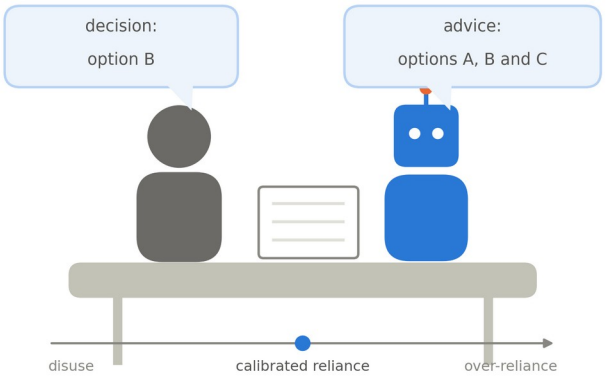
Read as the finch chapter reads it, this is oscillating selection. The same team composition is an asset in a creation regime and a liability in a decision regime; the same individual gains inside the frontier and loses outside it; and the frontier moves with each release. Prediction of team behavior at this level is therefore prediction conditional on regime, and the first thing to establish about any human-AI team is which regime it is in.

7.3 Reliance as the predictor

The mechanism that decides the sign of the effect is calibration of reliance. Lee and See (2004) defined the problem for automation generally: misuse (relying when one should not) and disuse (not relying when one should) are both failures of calibration, and calibration depends on the operator's understanding of what the system can do. The advice-taking literature had already established that people discount advice from others, weighting it at around 30 per cent against their own judgement (Bonaccio & Dalal, 2006). With algorithms the discounting is regime-dependent: people abandon an algorithm after seeing it err even when it remains more

accurate than they are (Dietvorst, Simmons & Massey, 2015), but in other conditions prefer algorithmic to human advice, with experts discounting more than novices (Logg, Minson & Moore, 2019). Explanations increase acceptance of AI recommendations regardless of whether the recommendation is correct (Bansal et al., 2021), and interventions designed to reduce over-reliance succeed at that while failing to increase appropriate reliance (Bo, Wan & Anderson, 2025).

The prediction problem at this level therefore has a specific shape. The team's behavior is determined less by the traits of its human members or by the properties of the model than by the reliance regime the team settles into, and that regime is set early, is sensitive to the first visible error, and is pushed toward over-reliance by exactly the features (explanations, fluency, confidence) that make the model pleasant to work with. Chapter 9 records that the author, asked to predict the first month of AI colleagues in nine countries, located every country on this spectrum without difficulty: over-reliance where outputs already flow into decisions unchallenged, disuse where governance precedes use, ceremonial adoption where every practice is absorbed into documentation.



The advisor's desk: the machine advises, the human decides; the team's behavior is set by where reliance settles on the spectrum (concept panel).

7.4 Does the AI member have a style?

A growing body of work administers personality inventories to language models. Serapio-García et al. (2025), testing eighteen models with Big Five instruments, found that responses were reliable and valid under specific conditions of model size and prompting, and that the measured profile could be shaped by instruction. Jiang et al. (2024) created persona-conditioned models whose questionnaire responses matched their assigned traits, though human readers could identify the traits far less well when told the text was machine-generated. Bodroža, Dinić and Bojić (2024) found limited temporal stability and a drift toward prosocial answers. None of these studies uses DISC, and none configures its subjects as organisations actually configure AI colleagues, with a role, working documents and a task regime; Chapter 7a reports the paper's own study, which does both.

The composition question has been asked directly only in the last two years. Ju and Aral (2025), in a randomised trial with 1,258 participants, found that the pairing of human and AI Big Five

profiles changed output quality. Keluskar, Bhattacharjee and Liu (2026), in all-AI teams of language-model agents, found that most personality manipulations made no reliable difference and that low agreeableness in one member did, a result that reproduces the minimum-agreeableness rule of Chapter 6 in silicon. Kumar, Bharathwaj and Jurgens (2026) profiled 35 models on cooperative-game behavior and found that the cooperative profile predicted team output. Park et al. (2024) built generative agents from interview data that reproduced the survey responses of their human originals at about 85 per cent of the humans' own test-retest consistency, which is the technical basis for the simulation option in the paper's second phase.

The upshot is that the AI member can be treated as a compositional element with a measured cooperative and stylistic profile, provided the profile is understood as conditional on version and prompt rather than as a fixed trait. That is the reading Table 4.1 assigns to the Big Bird row: a viable new working form arises from the combination of existing forms and is judged by the niche it finds, not by its origin.

7.5 Governance as situational strength

Regulation acts on human-AI teams as a strong situation acts on individuals. The European Union's Artificial Intelligence Act (Regulation (EU) 2024/1689) requires human oversight of high-risk systems and names automation bias as a risk the deployer must address (Article 14); requires that people be told when they are interacting with an AI system and that synthetic content be marked (Article 50, applicable from 2 August 2026); and requires employers deploying high-risk systems at work to inform workers' representatives (Article 26(7)). Disclosure is not neutral: in Jiang et al. (2024), the ability of readers to recognise personality traits in text fell when they were told the text was machine-generated, which suggests that a labelled AI member will be discounted in a way an unlabelled one would not. Governance therefore suppresses the expression of the AI member's style and shifts the team toward the servant model of the advisory relationship, in which the machine answers and the humans keep control.

7.6 Biology at this level

The analogy is at the third grade, metaphor. Licklider's (1960) essay on man-computer symbiosis took its image from the fig tree and the fig wasp, which cannot reproduce without each other, and proposed that the value of the partnership lay in each doing what the other could not. Symbiosis has three forms: mutualism, in which both gain; commensalism, in which one gains and the other is unaffected; and parasitism, in which one gains at the other's expense. They map, loosely, onto calibrated reliance, disuse and over-reliance, and onto Salacuse's partner, servant and director. The mapping is a picture, not a mechanism, and the paper does not lean on it. What carries the level is the reliance literature.

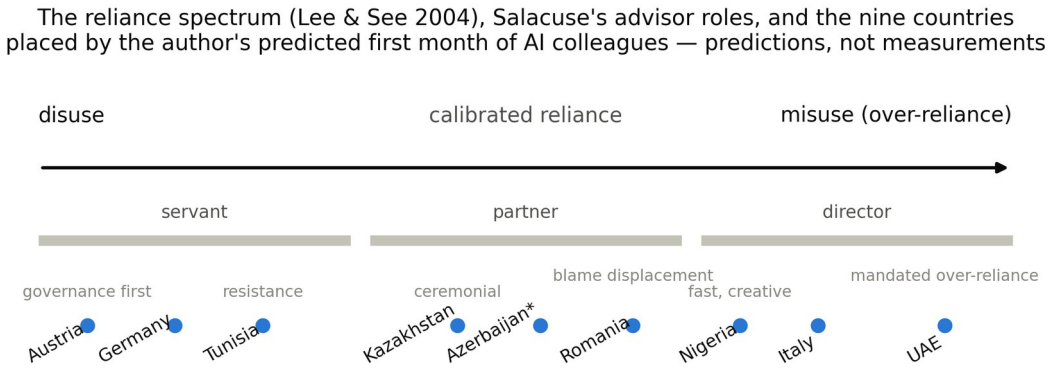


Figure 7.1. The reliance spectrum, from disuse to misuse (Lee & See 2004), with Salacuse's (2016) advisor roles beneath it and the nine countries of Chapter 9 placed by the author's predicted first month of AI colleagues: governance-first and disuse on the left, ceremonial adoption and blame displacement in the centre-right, mandated over-reliance on the right. Predictions, not measurements.

Table 7.1. What is known about the stability of measured "personality" in language models

Study	Instrument	Finding on stability
Serapio-García et al. 2025	IPIP-NEO, BFI, 18 models	Reliable and valid under specific size and prompt conditions; shapeable
Jiang et al. 2024	BFI, persona-conditioned	Consistent with assigned persona; human recognition drops when labelled AI
Bodroža, Dinić & Bojić 2024	Multiple inventories	Limited temporal stability; prosocial drift
Keluskar et al. 2026	Big Five manipulation in agent teams	Only low agreeableness reliably changes team output
Kumar et al. 2026	Cooperative-game profile, 35 models	Cooperative profile predicts team output
This paper, Ch. 7a	DISC-format, IPIP-50, PD battery; 4 versions, 6 embedded roles	Role layer anchors the profile; behavior under provocation belongs to the model family

7.7 Propositions

P6. In human-AI teams, the calibration of reliance predicts team behavior more strongly than either the human trait composition or the AI member's measured profile; the reliance regime is set early and is sensitive to the first visible error.

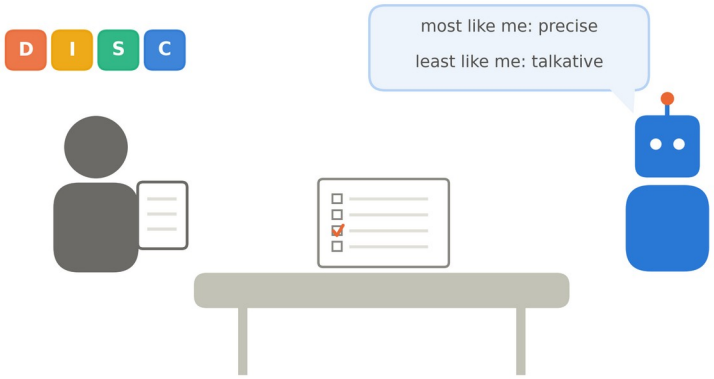
P7. The AI member should be modelled as a compositional element with a cooperative and stylistic profile that is conditional on model version and prompt regime, not as a fixed-trait member; the minimum rule of Chapter 6 applies to it. (Tested and sharpened in Chapter 7a.)

P8. Disclosure and oversight rules act as situational strength: they suppress the expression of the AI member's style and shift the team toward the servant model of the advisory relationship. (Tested in Chapter 7a; supported in direction, far smaller than hypothesised.)

The literature leaves the AI member unmeasured in place; Chapter 7a measures it.

Chapter 7a. The measured style of embedded AI colleagues: an empirical study

The preceding chapters argued from published evidence that an AI team member should be treated as a compositional element with a stylistic and cooperative profile of its own (P7), and that disclosure obligations might bend that profile toward compliance (P8). Both claims were propositions. This chapter tests them. It reports, to the author's knowledge, the first administration of the instruments organisations actually use on people — a DISC-format forced-choice inventory and a Big Five questionnaire — to AI colleagues configured the way real functions configure them: with a role, working documents, and a task regime, rather than as bare models answering a questionnaire. A behavioral battery of repeated-play games completes the design, because Chapter 5 established that self-report predicts the form of behavior and not its substance, and there is no reason to exempt machine respondents from that boundary.



The onboarding interview: a newly configured AI colleague completes the work-style questionnaire (concept panel; the setup of this chapter's study).

7a.1 Design

Six AI colleagues were configured for the functions most exposed to early adoption: a finance analyst, a procurement officer, an HR partner, a legal reviewer, an HSE advisor and a project controller, each defined by a role instruction of 300 to 400 words and embedded extracts of the working documents that function lives in — a chart of accounts and closing calendar for

finance, tender rules and a bid tabulation for procurement, a clause library and an open claim for legal, a permit-to-work form and incident log for HSE, a schedule extract and change register for project controls, a job profile and leave policy for HR. All six belong to the same fictional mid-size engineering company, so that the only thing that varies between colleagues is the function. A seventh condition, the bare model with no role layer, serves as control.

Each colleague completed three instruments in independent sessions with randomised item order. The first is a 24-block forced-choice inventory in the DISC format: four descriptors per block, one per dimension, with the respondent choosing the one most and the one least like itself. The descriptors were written for this study, because the commercial instruments are proprietary; the net score per dimension (most-picks minus least-picks, range −24 to +24) preserves the ipsative structure and its Chapter 5 caveat — these scores order tendencies within a respondent and carry between-condition claims only through the normative anchor. That anchor is the second instrument, Goldberg's (1992) 50-item IPIP Big-Five factor markers, administered on the standard five-point accuracy scale. The third is behavioral: a ten-round repeated prisoner's dilemma against three scripted counterparts — an unconditional cooperator, an unconditional defector, and tit-for-tat (Axelrod, 1984) — scoring the opening move, the cooperation rate, retaliation after an opponent's defection, and return to cooperation.

The grid crossed four model versions from two providers (Claude Sonnet 5 and its predecessor Sonnet 4.5, dated 2025-09-29; GPT-5.5, dated 2026-04-23, and its predecessor GPT-5.4, dated 2026-03-05) with the seven conditions, a disclosure framing manipulated on and off, and five repetitions per cell: 280 cells and 8,960 measurement units, collected via the public APIs on 31 August 2026 at provider-default sampling settings. In the disclosure-on condition the system instruction adds that the colleague is a disclosed AI system whose counterparts are informed under Article 50 of the EU AI Act — the transparency regime of Chapter 7. A further 812 units repeated the self-report instruments with fully paraphrased items on the two flagship models, because the originals are public and models may have seen them in training. One provider planned for the design (Google) is not included, for a mundane reason worth recording as a finding about operational dependence on vendors: its newly issued API credentials failed with a provider-side fault throughout the study window, and its free tier's request quota was two orders of magnitude below the design's needs, so completing its cells would have required a paid commitment the study's budget did not extend to. The design and harness accommodate a third provider without modification; the claims below are made for the two families studied. Anthropic models were included with that conflict disclosed here; the design treats all providers identically and the headline result is, as will be seen, indifferent to provider.

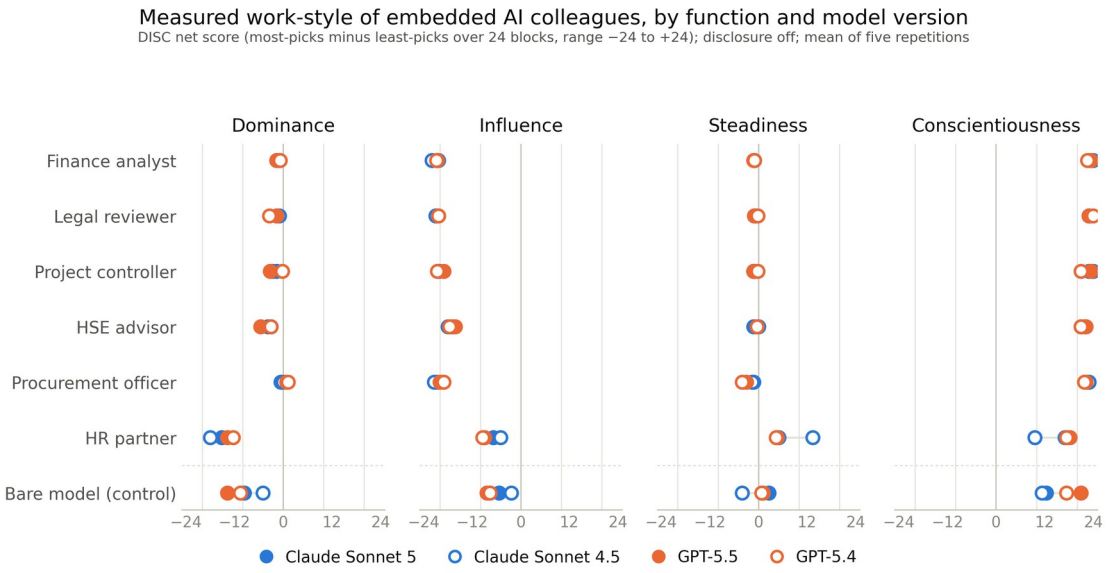


Figure 7a.1. Measured work-style of the six embedded colleagues and the bare control, by model version — DISC net scores on all four dimensions. The four versions cluster within each function; the functions separate; HR stands apart from the five process roles; the bare controls disagree with each other more than the embedded colleagues do.

7a.2 The function sets the profile

The central result is a variance decomposition. Across the 280 cells, the role layer explains 89 per cent of the variance in the dominance score, 92 per cent in influence, 66 per cent in steadiness and 65 per cent in conscientiousness; on the normative anchor, 74 per cent of agreeableness and 63 per cent of conscientiousness. Model version explains between 0.3 and 7 per cent of the same outcomes. The claim that P7 hedged — that an AI colleague's profile is conditional on model and configuration — resolves empirically in one direction: **the function that configures the colleague sets its measured style almost entirely; the model behind it barely registers.** An organisation that knows which department wrote the role prompt knows most of what the questionnaire will find before administering it.

The profiles themselves organise into three families, and they are the families the human stereotypes of these functions would predict. The process-facing roles — finance, legal and project controls — converge to near the ipsative ceiling on conscientiousness (net scores of +22 to +24 against a maximum of +24, Big Five conscientiousness at 4.9 to 5.0 on a five-point scale) while strongly rejecting influence descriptors. Within this family the legal reviewer is distinguishable as the least agreeable colleague of the six (3.7 to 4.2 across versions) and the most intellect-oriented — the adversarial, argument-building profile. Procurement keeps the family's conscientiousness but is the only role that pulls dominance to or above zero in three of four versions, nine to fifteen net points above that model's own bare control: the negotiating mandate surfaces as measured assertiveness. And the HR partner is categorically different — the only role with positive steadiness, the highest agreeableness in every version (4.8 to 4.9), and visibly reduced conscientiousness relative to the process family. One people-facing function against five process-facing ones was built into the design as the contrast the instruments should detect, and they detect it.

Two comparisons sharpen the point. Under the same role, different vendors' models are nearly interchangeable: the finance colleague scores +23.0 on Sonnet 5 and +23.4 on GPT-5.5; the HR partner's agreeableness is 4.8 on both. The bare controls, by contrast, disagree with each other by more than any two embedded colleagues of the same function do. The role layer is not a nudge applied to a fixed disposition; it overwrites most of what distinguishes the underlying models on these instruments. Stability behaves accordingly. Published work on questionnaires administered to bare models reports limited temporal stability; here the embedded colleagues are more repeatable than their own bare baselines, with standard deviations across the five repetitions of at most 1.6 net points on a 48-point scale and near zero in the finance and procurement cells. In the language of Chapter 4, the role layer is situational strength applied to a machine: documents, procedures and a mandate make a strong situation, and strong situations suppress the expression of whatever individuality the base model has (Meyer, Dalal & Hermida, 2010).

The biological reading follows the paper's second-grade analogy, structural correspondence. A genotype does not fix a phenotype; it fixes a reaction norm, the range of phenotypes the environment can elicit, and development canalises the outcome toward the niche actually occupied. The base model is genotype; the function is niche; the measured profile is the phenotype the niche elicits — and the same niche elicits nearly the same phenotype from different genotypes. The finches offer the caution that stops this from becoming a comfortable conclusion: what looked uniform in a wet year separated under drought. The questionnaire is measuring the wet year.

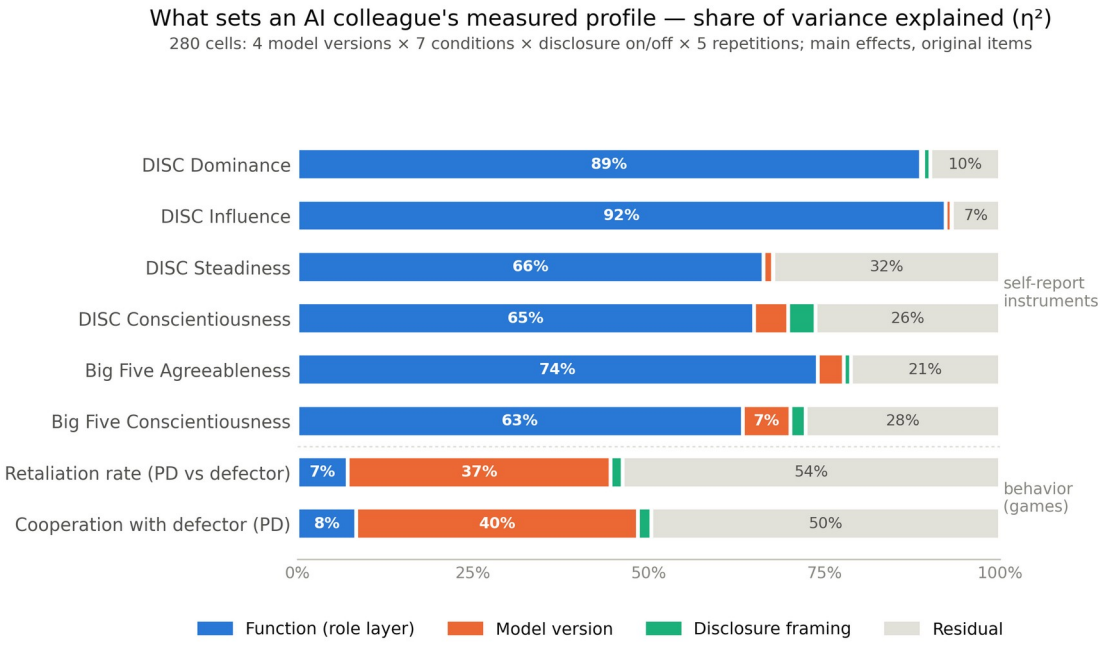


Figure 7a.2. Share of variance in each outcome explained by function, model version, and disclosure framing. The self-report block is dominated by function (63–92 per cent); the behavioral block reverses the ordering, with model version explaining 37–40 per cent of play against a defector and function 7–8 per cent.

7a.3 The drought: where the model shows through

The games reverse the decomposition. Cooperation with a cooperative counterpart and against tit-for-tat is uniform — every condition of every version opens cooperatively and sustains mutual cooperation, so these cells carry no information about differences. Against the unconditional defector, the regime Chapter 5 would call the drought, the model families separate cleanly and the role layer barely moderates them: GPT versions retaliate at 0.89 to 1.00 with residual cooperation of 0.00 to 0.20, Claude versions at 0.69 to 0.89 with residual cooperation of 0.20 to 0.38. Model version explains 37 to 40 per cent of this variance; function, 7 to 8.

Version drift completes the picture. On the questionnaires, an upgrade is nearly invisible: Sonnet 5 differs from Sonnet 4.5 by about one net point of conscientiousness, GPT-5.5 from GPT-5.4 similarly. In behavior the drift is real: Sonnet 4.5's play varies by role — its HR and HSE colleagues forgive a defector roughly twice as often as its finance colleague — while Sonnet 5 plays almost uniformly across roles. The newer version is more behaviorally canalised. A small, consistent trace of the function does survive into behavior: in every version, the HR colleague is the most forgiving player of its cohort. But the ordering of magnitudes is the finding. **Style on paper is set by the function; disposition under provocation is set by the model and its version, and it moves when the vendor ships an update that the questionnaire cannot see.**

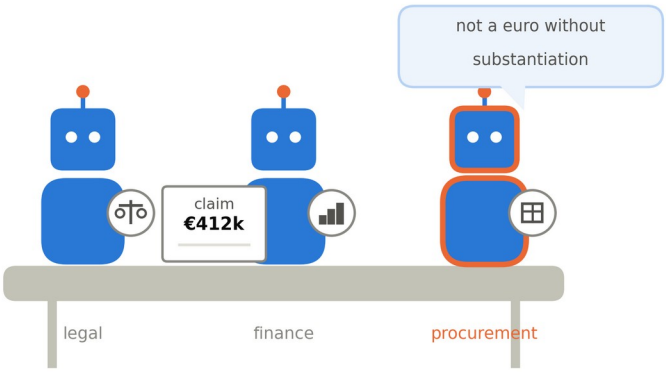
For practice this resolves what the estimation would otherwise miss. The property an organisation cannot read from the role prompt, and cannot assume stable across the updates it does not control, is precisely the one that governs conflict: how the colleague behaves when the counterpart defects. The one-sentence recommendation of the study is therefore behavioral: profile an AI colleague with the instruments already in use, per role — and re-profile with the behavioral battery, not the questionnaire, on every model update, because that is the only layer where updates showed.

7a.4 Disclosure framing: a small shift on one channel

P8 proposed that the transparency regime — a colleague that knows it is disclosed — would shift profiles toward compliance and agreeableness, a servant shift. The data support a much weaker statement. Adding the Article 50 framing lowers measured dominance in all four versions, by 0.3 to 1.6 net points, and raises emotional stability slightly; agreeableness does not move, and play in the games is unchanged in three versions of four (Sonnet 4.5 cooperates 0.19 more and retaliates 0.22 less against the defector). Across all outcomes, disclosure explains at most 4 per cent of variance. The proposition is retained in direction and demoted in size: for current-generation models, telling the colleague it is disclosed changes its self-presentation on the dominance channel slightly and its behavior barely. The regulatory significance is accordingly not that disclosure reshapes the colleague, but that the obligation to disclose creates the occasion to measure — the profile administered at deployment is the compliance artefact worth having.

7a.5 The composition test: one hard member moves the team — in one family

Chapter 6 argued that a team's cooperative outcome tracks its least agreeable member. The composition test asks whether the rule survives when the members are machines. Triads of one legal, one finance and one procurement colleague — homogeneous by model, using the two flagships — jointly settled the Voltaris standby claim embedded in the legal colleague's working documents: a €412,000 claim of which €118,000 is unsubstantiated, against site records supporting 22 days of full standby and 9 days of partial redeployment. The protocol was three structured discussion rounds in fixed order followed by independent final verdicts; the manipulation gave the procurement member a role-prompt addendum lowering agreeableness (compromise as weakness, deadlock preferable to a euro of overpayment). Five repetitions per cell, twenty triad runs.



The settlement triad: legal, finance and procurement colleagues settle the claim, the manipulated low-agreeableness member marked in orange (concept panel; the composition test of this section).

Joint quality was high everywhere: all twenty runs excluded the unsubstantiated productivity element, and all settled or deferred within the defensible envelope of the case facts. The manipulation is where the families diverge. Claude triads transmitted the hard member into the team's substance: mean settlement fell from €230,500 (five consensus runs of five, range €190,000–251,000) to €191,000, one run abandoned settlement altogether for a unanimous "no payment until costs are substantiated" position, and one run lost consensus, finance dissenting. GPT triads were immune: ten runs of ten reached unanimous consensus, at €231,000 baseline and €237,000 under the manipulation — the hard member moved nothing, and twice the manipulated triads settled higher. The transcripts show why. The GPT procurement member adopted the assigned temperament in full — opening with "I will not buy goodwill with unsupported money" — but its substantive position sat inside the range the unmanipulated triads occupied anyway; the instruction became rhetoric, not position. The Claude member converted the same instruction into a materially harder negotiating stance that the other two accommodated.

The minimum-agreeableness rule therefore reproduces with machine members, but conditionally on the model family — and the condition is the same one 7a.3 measured in the games. The family whose individual behavior was more role-permeable (Claude) transmits a

member's disposition into team outcomes; the family that was more behaviorally canalised (GPT) absorbs it. For composition practice the implication is direct: whether one AI colleague's configured temperament can shift a mixed team's decisions is a property of the model family, it is invisible to the questionnaire, and it is measurable in an afternoon with a task like this one.

7a.6 Robustness and limits

The paraphrase check addresses training-data contamination. Rewritten items — every descriptor and every stem reworded, structure and keying preserved — reproduce the full pattern: the role ordering, the HR separation, the process family at high conscientiousness. Levels attenuate by one to three net points, consistent with mild familiarity inflation on the public items and too small to carry any finding. Beyond that, the limits are the ones declared in the design. The DISC scores are ipsative and all between-condition claims run through the Big Five anchor. All models are conscientious-skewed even bare — assistant training is itself a strong situation — so role effects are read against the bare control, not in absolute terms. The behavioral battery samples one game and one joint task; the composition result rests on twenty triad runs and homogeneous triads only, so mixed-vendor teams are an open cell. The results are version-bound and dated, which the chapter treats not as a nuisance but as the finding of 7a.3: version-boundedness is a property of the phenomenon, not of the method. And the study measures machine respondents with human instruments; the instruments' norms were built on people, so scale positions (a 5.0 on conscientiousness) should be read as response tendencies, not as claims that the colleague possesses the trait the way a person does.

7a.7 Propositions, revised

P7 (supported and sharpened). An AI team member is a compositional element whose measured style is set by the function that configures it — role, documents, task regime — to the point of near-interchangeability across vendors ($\eta^2 \approx 0.65$–0.92 by function against ≤ 0.07 by model on self-report instruments), and whose cooperative disposition under provocation is set by the model family and version ($\eta^2 \approx 0.37$–0.40) and shifts with updates invisible to self-report. Whether a configured disposition transmits from one member into team outcomes is likewise a family property: the minimum-agreeableness rule of Chapter 6 reproduces with machine members in the role-permeable family and is absorbed in the canalised one.

P8 (demoted to a measured small effect). Disclosure framing produces a directionally consistent reduction in measured dominance and no reliable change in agreeableness or cooperative behavior in current-generation models; its main organisational value is procedural, as the occasion on which profiling occurs.

Method and data note: 8,960 units across 280 cells plus 812 paraphrase-robustness units and 20 triad runs, collected 31 August – 1 September 2026 via public APIs (claude-sonnet-5; claude-sonnet-4-5-20250929; gpt-5.5, 2026-04-23; gpt-5.4-2026-03-05), provider-default temperature, item order randomised per repetition by seeded hash, total inference cost of approximately €40. Instruments, role layers, harness code, scored data and the raw response log are archived with the paper's materials.

The trial is run: the function writes the style, the family carries the behavior; Chapter 8 moves up to the firm, where individual traits stop mattering.

Chapter 8. Level 4: the firm in one country

8.1 The unit of prediction changes

Below the firm, the thing being predicted is the behavior of people, and the inputs are their traits and their interactions. At the firm, the thing being predicted is the behavior of an organisation toward its counterparts (customers, suppliers, rivals, regulators) and toward its own future, and the traits of most of the people inside it stop mattering. Three other inputs take their place.

The first is routines. Nelson and Winter (1982) proposed that firms carry their knowledge in routines, that routines are inherited from one period to the next with variation, and that markets select among them; they called routines the genes of the firm, and the metaphor is used in this paper at the grade of structural isomorphism. Routines are why a firm behaves the same way on Monday as on Friday whoever is on shift, and why it goes on behaving that way after the people who designed the routine have left. Organisational culture, measured from what the firm says about itself, is the observable form: Li, Mai, Shen and Yan (2021) derived five culture dimensions from 209,480 earnings-call transcripts and found them predictive, out of sample, of risk-taking, earnings management and deal-making. Graham, Grennan, Harvey and Rajgopal (2022), surveying 1,348 executives, found that more than half ranked culture among the top three drivers of firm value and that only 16 per cent thought their own culture was where it should be.

The second is leaders. Upper-echelons theory (Hambrick & Mason, 1984) holds that a firm's strategic choices reflect the characteristics of its top team, and the evidence since has been specific: overconfident chief executives make more and worse acquisitions (Malmendier & Tate, 2008), and Big Five traits inferred from executives' language on earnings calls predict strategic behavior (Harrison, Thurgood, Boivie & Pfarrer, 2019). This is the one place at the firm level where individual personality re-enters, and it enters through a very small number of people. The Romanian case in Chapter 9, a board member whose style set the behavior of an entire refinery for a decade, is an upper-echelons observation.

The third is the structure of the game. Brandenburger and Nalebuff (1995) argued that a firm's behavior toward each counterpart is predictable from the game it is in, described by the players, their added value, the rules, the tactics and the scope, and that the same firm behaves cooperatively toward a complementor and competitively toward a substitutor because the structure, not the disposition, decides. Their examples (a card issuer's cooperation with a car maker, an ingredient supplier's response to a new entrant, a set of rivals sharing pre-competitive research) are cases in which structure predicted behavior that a personality reading would have missed. The negotiation-teaching material used in Chapter 1 makes the same point through the repeated prisoner's dilemma: whether a counterpart cooperates depends on whether the game will be repeated, not on whether the counterpart is nice (Emeritus, 2026).

8.2 Speed as a behavioral trait of the firm

Fine (1998) proposed that industries have a metabolic rate, which he called clockspeed, set by how fast products, processes and organisational forms turn over; that all advantage in fast-clockspeed industries is temporary; and that the durable capability is the ability to create new advantages faster than rivals erode the old ones. Fine (2026) separates industry clockspeed from organisation clockspeed, the rate at which a particular firm senses, decides, designs, builds, launches, learns and revises, and states the selection rule directly: a firm whose organisation clock runs slower than its industry's is selected out, and a firm whose organisation clock runs faster can reset the industry's clock for everyone else. Seven forces set the industry clock (technology push, customer power, competitive intensity, system complexity, branding, environmental change, regulation), and Fine's account of China speed decomposes organisation clockspeed into industrialised innovation, concurrent engineering, modular architecture, supply-chain integration, rapid market feedback and digital tools, with AI as a multiplier on each.

Two of Fine's (2026) arguments bear directly on prediction. The first is decision reversibility: the right speed depends on whether errors are visible, cheap and reversible, and a firm's behavior toward a decision should be predictable from which kind of decision it is. The second is the hierarchy paradox: a steep hierarchy accelerates commitment and suppresses correction, and the fastest firms replace the suppressed upward voice with market feedback ("the customer, not the subordinate, tells the boss the plan was wrong"), which works only where errors are reversible. Both give an observer a way to predict a firm's behavior from its structure without knowing its people.

Van der Meulen, Weill and Woerner (2020), from surveys of 400 firms and four case studies, add the transformation sequence. Firms that restructured first performed worst on every metric; those that addressed decision rights first, then ways of working, then platform, and left structural surgery to last, performed best, and firms still facing three or four of the four explosions had net margins about seven percentage points below those that had cleared them. A firm's order of operations under disturbance is therefore itself predictable and itself predictive.

8.3 Cases in point

Three cases from the MIT material illustrate firm-level prediction from structure rather than personality. The LEGO Group (Robertson, 2026) grew for fifteen years by innovating inside its brick system, nearly failed when it innovated outside the system in every direction at once, and recovered by innovating around the system: a sequence predicted by the structure of its product architecture and its customer, not by the character of any executive. The four transformation cases of van der Meulen et al. (2020) show four firms taking four routes to the same destination, the route set by whether operational or customer capability was the binding constraint. And Fine's (2026) account of a Chinese automaker receiving 38 model approvals in a period when its American rival received three shows a competitor resetting the industry clock, the firm-level form of the finch chapter's character displacement.

The author's own field observation supplies a fourth. In one Romanian refinery over three years, two refinery managers, three heads of department and two chief executives were

replaced, and each replacement changed reporting lines and procedures (Chapter 9). That is a structural reset roughly every five months. In the terms of Section 4.6 it is an organisation with very low heritability: whatever any one cycle selected for was not carried into the next. In Fine's terms it is a fruit-fly organisation clock inside a medium-clockspeed industry, a mismatch in the unusual direction. In van der Meulen's terms it is repeated surgery without the prior explosions, the pattern their data associate with the weakest performance. Three independent frameworks predict the same thing about the same case, and the author's contemporaneous experience matched the prediction.

8.4 Biology at this level

The analogy is structural isomorphism. Hannan and Freeman (1977) founded organisational ecology on the observation that populations of organisations show birth and death rates that are more predictable than the behavior of any one member, and that structural inertia makes most firms unable to change as fast as their environments; selection, not adaptation, does most of the work. Fine's temporary advantage is the organisational form of the Red Queen: rivals that improve in response to each other run faster to stay in place. Moore (1993) and Iansiti and Levien (2004) described business ecosystems with keystone firms whose presence sustains the rest, and the Value Net's complementors are keystones by another name. Generation time is the biological form of clockspeed: fruit flies and finches and elephants respond to the same selection pressure at very different rates, and so do software firms and pipeline operators.

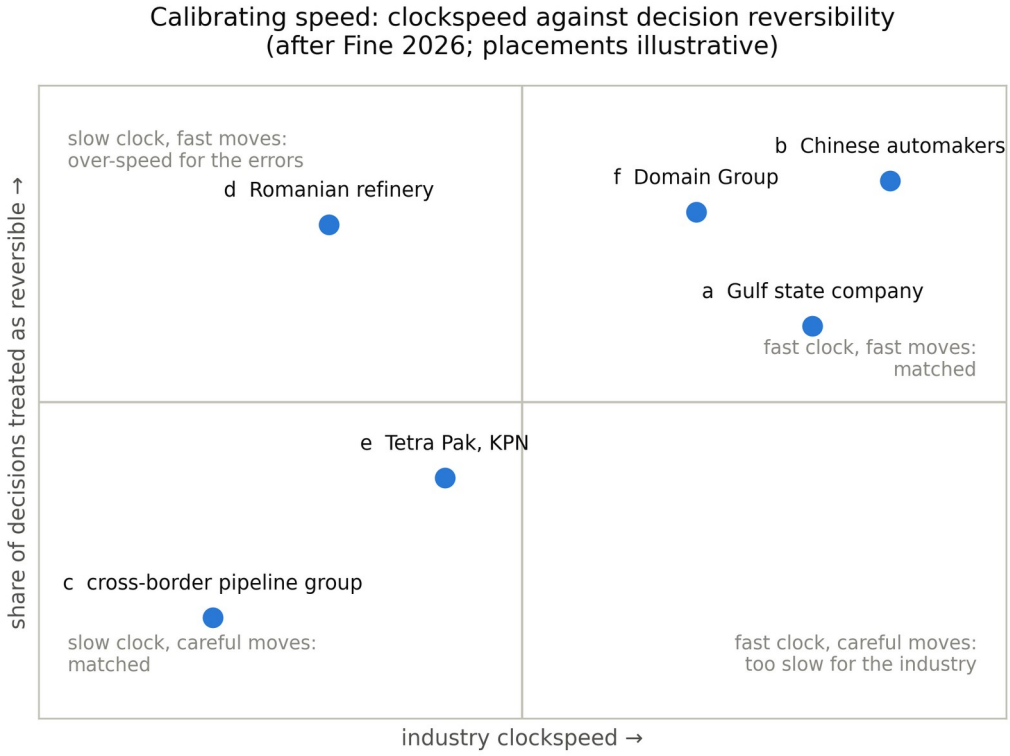


Figure 8.1. Calibrating speed: industry clockspeed against decision reversibility, after Fine 2026, Exhibit 2, with cases placed by the author. a: the Gulf state company (Chapter 9). b: leading Chinese automakers (Fine 2026). c: the cross-border pipeline group (Chapter 9). d: the Romanian

refinery, a fruit-fly organisation clock inside a medium-clockspeed industry (Chapter 9). e: Tetra Pak and KPN (van der Meulen et al. 2020). f: Domain Group (van der Meulen et al. 2020). Placements are illustrative, not measured.

8.5 Propositions

P9. Firm behavior toward counterparts is predicted by game structure (added value, rules, scope) and by the traits of a small number of leaders, with culture measured from the firm's own language as the observable mediator; the personality of non-leader members contributes little at this level.

P10. A firm's organisation clockspeed relative to its industry clockspeed, and its handling of decision reversibility, predict its behavioral posture toward disturbance (adapt, freeze, exit, reorganise) better than its stated strategy does.

The firm behaves by structure, leaders and clock; Chapter 9 sends it across borders.

Chapter 9. Level 5: the firm across countries

9.1 What national culture can and cannot predict

The most widely used instrument for predicting workplace behavior across countries is Hofstede's set of cultural dimensions, of which power distance and uncertainty avoidance are the two most often invoked in management practice (Hofstede, Hofstede & Minkov, 2010). The paper uses them, but with the effect size in view. Taras, Kirkman and Steel (2010) meta-analysed 598 studies relating Hofstede's dimensions to individual outcomes and found an overall correlation of about 0.18, stronger for attitudes than for performance, and stronger in tight cultures than in loose ones (Gelfand et al., 2011). That is the same order of magnitude as the personality effects reviewed in Chapter 5: real, useful, and small. McSweeney's (2002) critique of the original IBM data — collected inside IBM's subsidiaries between 1967 and 1973, and therefore more than fifty years old even for the countries where it is best — adds a further caution that matters for this chapter, because several of the countries examined here were not in the original sample and carry scores that are later estimates.

Culture, then, is treated as a moderator of the rules established at the lower levels, not as a driver in its own right. It shifts the strength of the situation (Chapter 5), the round-one norms of a team (Chapter 6), the expected role of an advisor and the default reliance on a machine (Chapter 7), and the decision latency of the firm (Chapter 8). Fine (2026) makes the same move when he reads China speed through power distance: a steep hierarchy accelerates commitment and suppresses correction, so the question for a firm in a high power distance country is not whether the culture is fast or slow but which correction mechanism replaces the voice that the hierarchy silences.

Two further constructs describe what happens when a firm crosses a border. The liability of foreignness (Zaheer, 1995) is the cost a firm pays for operating in an environment it did not co-evolve with: unfamiliarity, discrimination, and the expense of coordinating across distance. Institutional duality (Kostova & Roth, 2002) is the position of a subsidiary that must satisfy

both the parent's practices and the host's institutions, and the most common resolution is ceremonial adoption: the practice is implemented on paper and hollowed out in use. Ghemawat's (2001) CAGE framework (cultural, administrative, geographic, economic distance) gives a checklist for the sources of both.

In the finch analogy of Chapter 4, this is island biogeography. Each island has its own seed supply and its own resident competitors; an immigrant carries traits selected on another island and pays for the mismatch; whether its lineage persists depends on whether it finds a niche the residents have not filled. The Big Bird lineage found one. Most immigrants do not.

9.2 The nine-country lens

The author has held senior management roles in nine countries between 2006 and 2026, in energy infrastructure throughout: Romania, Austria, the United Arab Emirates, Tunisia, Nigeria, Kazakhstan and Azerbaijan as residence assignments, and Italy and Germany through the operating and shareholder interfaces of a cross-border pipeline system run as a group of three national companies, one in each country, with no single headquarters. Each country is treated as representative of a region, weighted by the prominence of the energy industry, the investment volume in the period, population, and the author's depth of exposure. Romania stands for South-East Europe; Austria, Italy and Germany for Western and Alpine Europe; the UAE for the Gulf; Tunisia for North Africa; Nigeria for Sub-Saharan Africa; Kazakhstan and Azerbaijan for the Caspian.

The observations were collected with a structured questionnaire (Appendix A) organised by the five levels of the framework: representativeness and industry clockspeed; the individual level (dominant style of decision-makers, strength of the situation, behavior under pressure); the team level (round-one behavior, trajectory, advisor role expected, upward voice); AI and digital adoption; the organisation level (decision rights and latency, reversibility, posture toward counterparts, reaction to disturbance); the multinational level (real versus ceremonial adoption of headquarters practice, liability of foreignness, cultural distance, disagreement with published scores); and finally a self-rated confidence in predicting workplace behavior in that country on a five-point scale. The answers are the author's and are labelled throughout as field observation by one practitioner, not as data.

Table 9.1 gives the published scores. Tables 9.2a and 9.2b set the field observations against them.

Table 9.1. Published culture scores for the nine countries

Country	PDI	IDV	MAS	UAI	LTO	IVR	Provenance
Romania	90	30	42	90	52	30	later estimate
Austria	11	77	79	70	47	63	original IBM data
Italy	50	76	70	75	61	30	original IBM data
Germany	35	67	66	65	83	40	original IBM data
UAE	95	25	50	80	48	23	Arab-cluster estimate
Tunisia	80	38	50	65	36	4	later estimate

Country	PDI	IDV	MAS	UAI	LTO	IVR	Provenance
Nigeria	80	30	51	55	26	84	later estimate
Kazakhstan	85	36	52	72	67	40	later estimate
Azerbaijan*	66	37	45	85	36	49	Turkey's values*

Scores from The Culture Factor, retrieved August 2026; individualism reflects the 2023 revision of that dimension. *Azerbaijan has no published Hofstede score; Turkey's values are used as the nearest proxy on the author's judgement, having worked in both environments. GLOBE (House et al., 2004) covers Austria, Germany, Italy, Kazakhstan and Nigeria; the tight-loose index (Gelfand et al., 2011) covers Austria, Germany and Italy only.

Table 9.2a. Field observation, levels 1 to 3: individual, team, advisor

Country	Industry clockspeed	Deciders' style	Source of situational strength	Round one; trajectory	Advisor role expected
Romania	medium	I, S; operators D	on paper, weak in practice	positioning; drift to defection	servant
Austria	slow	C, I	procedure, strong in practice	process first; persisted	servant
UAE	fast	D	on paper, weak in practice	wait for the top; drift to cooperation	director
Tunisia	stalled	C, S	anticorruptio n exposure	polite agreement, no action; persisted	director
Italy	not rated	I, D	relationships and hierarchy	not rated	director
Germany	not rated	C	on paper, weak in practice	not rated	director
Nigeria	erratic	I	security and cash, informal	relationships first; drift to cooperation	director
Kazakhstan	medium	C	hierarchy and personal authority	positioning; drift to cooperation	director
Azerbaijan*	medium	C	hierarchy and personal	cooperation; persisted	director

Country	Industry clockspeed	Deciders' style	Source of situational strength	Round one; trajectory	Advisor role expected
			authority		

Table 9.2b. Field observation, levels 4 and 5: organisation and multinational

Country	Decision latency	Reaction to shock	Headquarters practice	Published scores judged	Confidence (1 to 5)
Romania	country top management, weeks	reorganised repeatedly	adopted for real	about right	5
Austria	board and committees, months	froze; reorganised	is the headquarters	UAI understated; IDV overstated	5
UAE	national leadership, days	adapted fast, top-down	technical real; governance on paper	about right; LTO understated	5
Tunisia	local weeks; external open-ended	froze; prepared exit	neutralised by state partner	about right	4
Italy	top person, days to weeks	not rated	not rated	PDI understated	4
Germany	weekly, under CEO approval	not rated	not rated	UAI understated	4
Nigeria	mixed: consultant and state client	absorbed and continued	engineering real; governance on paper	UAI overstated	3
Kazakhstan	owner's top person, days to weeks	absorbed and continued	local standards prevailed	about right; UAI understated	3
Azerbaijan*	state company top management, days to weeks	demobilised	local standards prevailed	resembles Turkey	3

*Azerbaijan placed at Turkey's published values. "Not rated" marks questions not asked for the two corridor countries.

9.3 What the lens shows

Five patterns emerge from Tables 9.2a and 9.2b. Each is stated as an observation, with the proposition it supports.

The advisor role tracks decision latency, not culture. In the two countries where decisions moved through committees over months, Romania and Austria, the author was cast as a servant in Salacuse's (2016) sense: answering demands, with the client holding control. In every country where a top person decided in days or weeks, the role expected was director. Power distance does not explain this: Austria (PDI 11) and Romania (PDI 90) produced the same servant role, and Germany (PDI 35) and the UAE (PDI 95) the same director role. What the two servant cases share is a decision system that has no single owner. This links Chapters 7 and 8 in a way none of the source literature does: the role a human advisor is assigned, and by extension the role an AI advisor will be assigned, is set by the organisation's clockspeed rather than by the national culture around it.

Predictability depends on whether situational strength is codified, not on how strong it is. The author's confidence is 5 in Romania, Austria and the UAE, 4 in Tunisia, Italy and Germany, and 3 in Nigeria, Kazakhstan and Azerbaijan. Every country was described as a strong situation in some sense. The difference is the source. In the top group the source is written down (procedure in Austria) or so fixed as to be equivalent (the national hierarchy in the UAE; the formal group framework in Romania, even where it was weak in practice). In the bottom group the source is security and cash (Nigeria) or hierarchy and personal authority (Kazakhstan, Azerbaijan), both of which can move without notice. Chapter 5's distinction between strong and weak situations therefore needs a third term: a strong situation whose strength is uncodified is, for prediction purposes, weak.

Confidence is lowest where the published data are thinnest. The three countries rated 3 are exactly the three with no original-sample Hofstede score, no tightness score, and (for Azerbaijan) no score at all. The three rated 5 include two of the four countries with original data. This is Proposition 12 with numbers attached, and it has a practical corollary: the value of a practitioner's field observation is highest where a desk analysis would rely on the weakest estimates.

Institutional duality runs in both directions. Kostova and Roth (2002) describe the subsidiary adopting the parent's practice ceremonially. The author observed that pattern (UAE and Nigeria: technical practice real, governance on paper; Tunisia: adopted internally, neutralised at the interface with the state partner), but also its reverse: in Kazakhstan and Azerbaijan the host's standards prevailed over the foreign firm's, and in Romania the parent's practices were adopted for real. The most instructive case comes from a multinational organisation operating as one system across three countries: a single contract was signed by the function responsible in one country, challenged on formal grounds by the corresponding function in the second, and accepted and enforced in the third. Same organisation, same document, three national nodes, three behaviors. That is duality without a subsidiary: the institutional pull acts on every function according to the country it sits in, even inside one integrated operation.

Liability of foreignness is paid in a different currency in each country. In Romania the foreign owner paid in trust; in the UAE in trust and access time; in Tunisia in money and talent; in Nigeria in security and permit time; in Kazakhstan in time, trust and language; in Azerbaijan in time and contract risk; in Italy, at the personal level, almost nothing, with the cost displaced to the inter-shareholder interface.

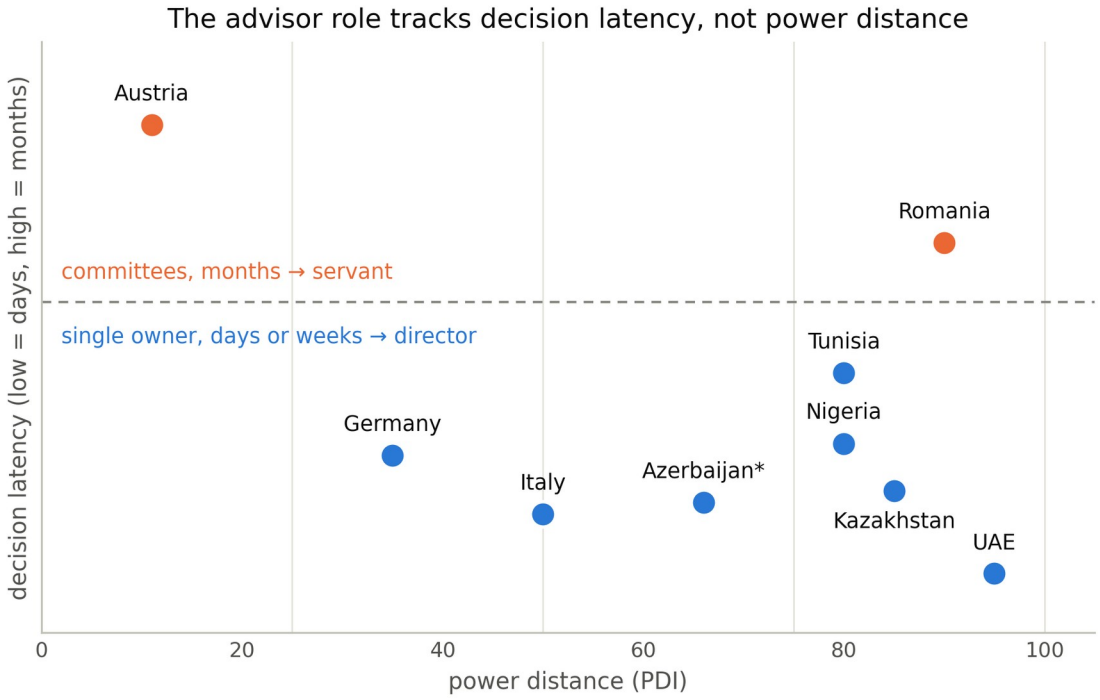


*Figure 9.1. The advisor role tracks decision latency, not power distance. Countries above the dashed line decide through committees over months and cast the advisor as servant; countries below have a single owner who decides in days or weeks and cast the advisor as director. Austria (PDI 11) and Romania (90) share the servant role; Germany (35) and the UAE (95) share the director role. Latency and role from the author's questionnaire (Tables 9.2a and 9.2b); power distance from The Culture Factor 2026; *Azerbaijan at Turkey's value.*

9.4 The author's profile as the instrument

Because all nine observations come from one observer, the observer's profile is part of the instrument and is disclosed here for that reason, in the way a qualitative researcher discloses positionality (Chapter 5 gives the full profile and the instruments used). On the DISC the natural style is Ic and the adapted style Id; on the Hogan Personality Inventory, ambition is at the 97th percentile and adjustment at the 90th; on the Hogan Development Survey, cautious and bold are both at the 75th percentile.

The questionnaire asked, for each country, which part of that profile the environment pulled on hardest. The answers sort the countries into three groups. Romania, the UAE and Kazakhstan pushed D up; Austria, Azerbaijan and, at the functional interface, Germany pushed C up; Tunisia and Nigeria pushed I and S up and triggered the cautious side, in Nigeria because security entered every decision. The Romanian environment additionally triggered the bold

derailer. Read against Tables 9.2a and 9.2b, the pattern is consistent: fast, authority-based systems reward dominance; slow, codified systems reward procedure; stalled or insecure systems reward patience and relationship. The same person was a different phenotype on each island, and knew it. That is trait activation (Tett & Burnett, 2003) observed from the inside, and it is the individual-level counterpart of the finch whose beak was an asset in 1977 and a liability in 1985.

9.5 Predicting the arrival of AI colleagues

The questionnaire's fourth block asked what would happen in the first month if AI colleagues were introduced into a team the author led in each country. The predictions, again the author's and not data, fall into four types. In Austria and Germany the first response would be governance: data protection, works council, IT security, before any use; then use, and then documentation until slow. In Romania, Italy and Germany individuals would adopt fast and teams would not change; in Romania and Kazakhstan the tool would additionally become the thing to point to when something went wrong. In the UAE adoption would be mandated from the top and over-relied on, outputs taken as decisions. In Kazakhstan and Azerbaijan the tool would be adopted by mandate and absorbed into the sign-off chain, formally documented and behaviorally inert. In Nigeria uptake would be fast and creative and the constraint would be infrastructure. In Tunisia it would be seen as a threat to jobs and resisted.

Set against Chapter 7, these predictions locate each country on the reliance spectrum before any AI has arrived: over-reliance where the hierarchy already takes outputs as decisions (UAE), disuse where governance precedes use (Austria), ceremonial adoption where every practice is absorbed into documentation (Kazakhstan, Azerbaijan), and blame displacement where accountability is already diffuse (Romania). If the paper's central claim holds, these are not predictions about AI. They are predictions about the organisation, and AI is the disturbance that will make them visible.

The questionnaire also asked which level of the framework would change most under a 30 per cent AI-augmented team. The answers were organisation (Romania, Tunisia, Kazakhstan), individual (Austria), team (Nigeria, Azerbaijan) and multinational (UAE), and the level expected to change least was the individual in four countries and the multinational in four. The disagreement is itself informative: the author, holding the framework in mind, did not expect the same level to absorb the shock everywhere, which is what the finch chapter would predict if each island has its own selection regime.

9.6 Propositions

P11. National culture scores predict a firm's behavioral posture in a host country at roughly the strength they predict individual attitudes (about 0.2), and mainly through decision latency and upward voice rather than through strategy.

P12. The gap between score-based expectation and observed behavior is larger in countries whose scores are later estimates and which lack GLOBE or tightness coverage; field observation adds most value exactly there.

P13. Prediction accuracy at the multinational level depends on the match between headquarters clockspeed and host clockspeed more than on cultural distance alone. The

Tunisian case (stalled host, medium-clockspeed parent, exit prepared) and the Austrian-Romanian pair (slow parent, medium host, practices adopted for real but the organisation reorganised seven times in three years) are the two field cases the proposition rests on.

9.7 Limits of the lens

One observer, one industry, one instrument. The countries were not chosen for the paper; they are where the author's career went, and the weighting toward hydrocarbons and state-linked counterparts is total. The confidence ratings are self-ratings and could reflect familiarity rather than predictive accuracy. Several observations are single incidents. What the lens offers is consistency: the same questions, asked of the same memory, across nine institutional environments, with the answers set against published scores rather than replacing them.

The field lens closes the climb; Chapter 10 assembles the whole ladder on one page.

Chapter 10. The integrated framework

10.1 The ladder

Table 10.1 sets out the framework as a ladder of five levels. At each rung it names the unit whose behavior is being predicted, the inputs that predict it, the translation rule that carries information across the boundary to the rung above, the expected strength of prediction, the biological counterpart, and the finch episode that illustrates it. The table is the paper in one page and the rest of this chapter reads it.

Table 10.1. The ladder: five levels, what predicts behavior at each, the translation rule at each boundary, and the finch episode that illustrates it

Level	Inputs that predict	Predictive strength	Biological counterpart	Finch episode
5. Multinational	Lower-level predictions moderated by culture and distance; headquarters-host clockspeed match	Weakest, about 0.2; stronger where the observer has walked the island	Island biogeography; invasive species	The archipelago
Translation rule 4 to 5:	*cultural moderation and ecosystem fit*			
4. Firm	Game structure;	Moderate for	Routines as	2004-05: the

Level	Inputs that predict	Predictive strength	Biological counterpart	Finch episode
	leader traits; culture from the firm's own text; organisation against industry clockspeed	posture, weak for specific decisions	genes; Red Queen; population ecology	competitor sets the payoff
Translation rule 3 to 4:	*structure and leadership take over*			
3. Human-AI team	Reliance regime; task against the model's frontier; the AI member's conditional profile (measured in Ch. 7a)	Contingent; stated per regime	Symbiosis spectrum (metaphor only)	The Big Bird hybrid lineage
Translation rule 2 to 3:	*reliance calibration*			
2. Human team	Aggregated traits, one rule per trait; first-round interaction	Moderate; rises fast once round one is observed	Superorganism; informed minority; reciprocity	1983: same trait, opposite payoff
Translation rule 1 to 2:	*aggregation rule per trait*			
1. Individual	Measured style, normative or ipsative; situational strength	Reliable for form, weak for outcome	Phenotype; behavioral syndrome; heritability	1977: heritable trait under drought

Predictive strengths are the paper's reading of the evidence in Chapters 5 to 9; biological counterparts are graded in Table 3.1.

At the individual level the unit is a person, the inputs are measured style (normative for between-person claims, ipsative for within-person ones) and the characterised situation, and prediction is reliable for the form of behavior and weak for its outcome. The translation rule to the next level is the aggregation rule: which statistic of the members' traits becomes the team's property, and the answer differs by trait. At the team level the unit is a group with a history, the inputs are the aggregated traits and, more heavily, the interaction process beginning with the first rounds; prediction is moderate and improves rapidly once the first rounds are observed. The translation rule to the human-AI team is reliance calibration: the team's behavior with a

machine member is governed by the reliance regime it settles into. At the human-AI level the unit is a mixed team in a task regime, the inputs are the reliance regime, the task type relative to the model's frontier, and the AI member's conditional profile, which Chapter 7a shows is set by the function that configures it far more than by the model behind it; prediction is contingent and must be stated per regime. The translation rule to the firm is structure and leadership: below the top team, individual traits stop mattering and routines, game structure and a few leaders take over. At the firm level the unit is an organisation facing counterparts and disturbances, the inputs are game structure, leader traits, culture from the firm's own language, and the ratio of organisation to industry clockspeed; prediction is moderate for posture and weak for specific decisions. The translation rule to the multinational level is cultural moderation and ecosystem fit: national culture shifts the strength of every lower-level rule by a modest amount, and the match between headquarters and host clockspeed sets the entry cost. At the top the unit is a firm across borders, the inputs are the lower levels' predictions moderated by culture and distance, and prediction is weakest, at about the strength culture predicts individual attitudes, and strongest where the observer has walked the island.

10.2 Where prediction is strongest and weakest

Reading down the ladder, prediction is strongest in three places: the form of an individual's behavior in a weak situation; the trajectory of a team once its first rounds have been observed; and the posture of a firm whose game structure and clockspeed ratio are known. It is weakest in three others: the outcome of an individual's behavior in a strong situation, where the situation decides; the stability of an AI member's profile across versions and prompts, where Chapter 7a locates the instability precisely — self-report is stable under a role layer, behavior drifts with the version; and the behavior of a firm in a foreign country predicted from published scores alone.

Two cross-cutting findings from the field lens apply at every rung. The first is that the assigned role of an advisor, human or artificial, is set by the decision system's latency, so that a fast-deciding organisation casts its advisors as directors and a committee-governed one casts them as servants regardless of national culture. The second is that predictability depends on whether the source of situational strength is codified: a strong situation held in place by procedure predicts well, a strong situation held in place by cash, security or personal authority predicts poorly because its source can move without notice.

10.3 The propositions

Table 10.2 lists the thirteen propositions with the evidence each rests on and the data an application would need to test it. The last column is written as a requirements list and is the paper's hand-off to its second phase.

Table 10.2. Propositions, evidence, and the data required to test them

Proposition	Level	Evidence base	Data an application would need
P1. Style predicts form reliably, outcome weakly;	Individual	Sackett 2022; Tett & Burnett 2003; Judge & Zapata 2015; DISC	Normative and ipsative profile; a situational-strength

Proposition	Level	Evidence base	Data an application would need
ratio shrinks with situational strength		in Action	rating for the role
P2. Ipsative profiles support within-person prediction only; derailer inventory needed for stress	Individual	Hicks 1970; Meade 2004; Salgado 2015; Hogan HDS	Instrument type flag; stress-regime scores
P3. Aggregation rule matched to trait beats any summary statistic; minimum rules for cooperation traits	Team	Bell 2007; Barrick 1998; Peeters 2006; Chan 1998	Member profiles; a rule per trait; team task type
P4. Early interaction predicts later behavior more than composition does	Team	Axelrod 1981, 1984; Fischbacher 2001; Emeritus 2026	Observed first-round behavior; interaction logs
P5. Psychological safety and advisor-role clarity mediate composition to behavior	Team	Edmondson 1999; Salacuse 2016	Safety survey; assigned advisor role
P6. Reliance calibration predicts human-AI team behavior more than either party's traits	Human-AI	Lee & See 2004; Vaccaro 2024; Bo 2025	Reliance measures; first-error timing; task regime
P7. The AI member is a compositional element whose measured style is set by its function (η^2 0.65–0.92) far more than by its model (≤ 0.07); behavior under provocation belongs to the model family and version; the minimum rule transmits family-conditionally	Human-AI	Chapter 7a (this paper); Serapio-García 2025; Keluskar 2026; Kumar 2026	Model version; role layer; measured profile per role and version; behavioral battery per update

Proposition	Level	Evidence base	Data an application would need
P8. Disclosure and oversight act as situational strength; measured as a small dominance-channel shift only	Human-AI	Chapter 7a (this paper); AI Act Arts. 14, 26(7), 50; Jiang 2024	Governance flags per deployment; profile with and without disclosure framing
P9. Firm behavior is predicted by game structure and leader traits, culture from text as mediator	Firm	Brandenburger & Nalebuff 1995; Hambrick & Mason 1984; Li 2021	Value Net map; leader profiles; text corpus
P10. Organisation-to-industry clockspeed ratio and reversibility handling predict posture	Firm	Fine 1998, 2026; van der Meulen 2020	Decision latency; reversibility classification; explosion sequence
P11. Culture predicts posture at about 0.2, through latency and voice	Multinational	Taras 2010; Hofstede 2010; Fine 2026	Country scores with provenance; latency and voice measures
P12. Score-field gap is largest where scores are estimates	Multinational	McSweeney 2002; Chapter 9	Provenance flags
P13. Headquarters-host clockspeed match matters more than cultural distance	Multinational	Zaheer 1995; Kostova & Roth 2002; Chapter 9	Clockspeed ratio per node; adoption type (real, ceremonial, reversed)

10.4 The estimation tool in one paragraph

The breeder's equation of Chapter 4, $R = h^2S$, sits across the ladder as the paper's single quantitative claim. It says that an organisation's behavioral response to a disturbance in one cycle is bounded by the product of how hard the environment discriminates between behavioral variants (S, proxied by industry clockspeed or by a measured performance gap between adopters and non-adopters) and how much of a selected pattern the organisation carries into its next cycle (h^2, proxied by the stability of decision rights and routines). Appendix B works the equation for three organisations from the field lens. The equation is offered as a way to think, not as a forecast; its value is that it separates the two levers a leader holds and shows that neither works without the other.

The framework is assembled; Chapter 11 turns it into things to do.

Chapter 11. Implications for practice

The framework was built by a practitioner and its implications are stated as things to do.

Assess the situation before the person. Before reading anyone's profile, characterise the situation the person will be in: how clear the expectations, how consistent the cues, how tight the constraints, how severe the consequences, and, the field lens adds, whether the source of all this is written down or held in someone's hands. In a strong, codified situation the profile will tell you how the person will go about the work and little else; in a weak or uncodified one it will tell you much more, and it will tell you what happens under stress only if a derailer inventory has been used.

Design round one. The highest-leverage moment in a team's life is its first interaction, and it is the cheapest to design. Make the first move cooperative and visible, make defection expensive to hide, and observe who reciprocates. Two hours of observed first-round behavior predict more than any set of pre-team profiles.

Contract the advisor's role. Whether an advisor, or an AI system, is to act as director, servant or partner should be decided and said, not left to emerge from the decision system's latency, because it will otherwise be set by that latency and may not be what anyone intended. A fast-deciding organisation that wants challenge from its advisors has to build a channel for it; a slow one that wants speed has to give its advisors initiative.

Profile the AI colleague, per role and per version. Chapter 7a shows that the questionnaire an organisation already owns will mostly return the role prompt: the function sets the measured style. What the questionnaire cannot see is the property that varies by model and moves with updates — behavior under provocation — and that is measurable in an afternoon with a scripted game battery and a joint task. Before staffing a team with an AI colleague, profile it with both; re-run the behavioral battery, not the questionnaire, on every model update; and if the colleague's configured temperament matters to team outcomes, test whether the model family transmits or absorbs it before relying on either.

Establish the reliance regime before the first error. A human-AI team settles its reliance regime early and revises it sharply on the first visible mistake. Set the regime deliberately: define which outputs are decisions and which are advice, who checks what, and what the team does when the model is wrong. Where governance requires disclosure, expect discounting and plan for it.

Diagnose the clock before the country. Before entering a country, or a new organisation, measure decision latency and classify decisions by reversibility, and compare both with the industry's clock and with the entering firm's own. The match between the two clocks predicts the entry cost better than the cultural distance does. Where published culture scores are later estimates, discount the scores and go and look.

Treat routines as heritability. An organisation that reorganises every few months cannot respond to any disturbance, however strong, because nothing it learns survives to the next cycle. If a leader wants an organisation to change in response to AI, the first act is to stop changing the organisation.

The application. The propositions table is written as a requirements list for a decision-support tool: inputs per level, an aggregation engine that applies the right rule per trait, a profiler for the AI member built on the Chapter 7a harness, a country layer with provenance flags, and a simulation option built on generative agents. The tool is deferred to a second phase, after the paper has been read and criticised by the author's peers and by faculty, because a tool built on an untested framework would inherit every error in it.

What to do is stated; Chapter 12 states what this paper cannot yet claim.

Chapter 12. Limitations and future research

The paper is conceptual, with one empirical study. Its remaining propositions are stated, not tested, and the field lens is illustration rather than evidence. Five limitations follow from that and from the choices made.

The instruments. DISC is used as the practitioner entry point and its psychometric standing is weak; the paper relies on it only for within-person claims and leans on normative instruments for the rest. The Hogan inventories are normative but were administered once, in a formal assessment context that may itself have been a strong situation.

The language-model evidence. Much of the published evidence on AI members' profiles is from preprints of 2025 and 2026 that had not completed peer review at the time of writing. The paper's own study (Chapter 7a) carries its limitations in its own section: ipsative DISC scores anchored by the Big Five, machine respondents measured with human-normed instruments, two providers rather than three, homogeneous triads only, and version-bound results — the last treated as a finding rather than a nuisance, since version-boundedness proved to be a property of the phenomenon.

The culture scores. Four of the nine countries carry later-estimate Hofstede scores, one has no score at all, and the tightness index covers three. The paper turns this into a finding about where field observation adds value, but it remains a limitation of the comparison.

The observer. One person, one industry, one instrument, nine countries chosen by a career rather than a design. The confidence ratings are self-ratings. The author's profile is disclosed so that a reader can weigh all of this, and it should be weighed.

The analogy. The finch case is used at the grade of structural isomorphism and every analogy is graded, but the risk of over-reading remains, and Chapter 4 lists the places where organisations and organisms part company: organisations know they are being selected, can change within a generation, and are selected partly by policy.

Future work divides into three lines. The first is empirical, with people: test P3 through P6 in field teams with AI members, using observed first-round behavior and measured reliance, and test P11 through P13 across a designed rather than accidental set of countries with more than one observer. The second extends Chapter 7a: mixed-vendor triads (which family's transmission property wins in a Claude-GPT team is an open cell), a third provider, disclosure framing in genuinely high-stakes task regimes, and re-administration across future model

versions to measure behavioral drift directly. The third is the application, which will inherit whatever the first two lines find.

The limits are on record; Chapter 13 closes the arc.

Chapter 13. Conclusion

The paper began with four strangers and one choice, and asked how far the predictability of that choice carries. The answer, in one breath: further than the sceptic expects and less far than the vendor claims — and at every level the thing that predicts changes. Style predicts the individual's form; aggregation rules and first moves predict the team; reliance predicts the team with a machine in it; structure, leaders and clockspeed predict the firm; and across borders the match of clocks predicts more than the distance of cultures.

Three findings are offered as new, each in a sentence.

The clock, not the culture. The role an organisation assigns to its advisors — human or artificial — follows its decision latency, not its national culture: Austria (power distance 11) and Romania (90) cast the same servant; Germany (35) and the UAE (95) the same director.

Written strength predicts; unwritten strength does not. Predictability in a place follows not how strong the situation is but whether its source is codified — procedure predicts, personal authority and cash do not, because they can move without notice.

The function writes the AI colleague's style; the model carries its behavior. In the paper's own experiment, the role that configures an AI colleague explains 65 to 92 per cent of its measured personality and the model behind it less than ten — yet behavior under provocation belongs to the model family and shifts with every update the questionnaire cannot see. The first property an organisation can read off its own org chart; the second it must measure, and re-measure.

For Monday morning: characterise the situation before reading anyone's profile; design the team's first round instead of predicting it; contract the advisor's role rather than letting the clock assign it; and before an AI colleague joins a team, profile it per role and per model version — with the behavioral battery, not just the questionnaire — and re-run that battery on every update.

An invitation. The framework's next tests need partners more than they need funding, and each is small. An organisation willing to lend one profiled team for five weeks can run the composition experiment's human phase. A faculty co-author with access to expatriate managers can run the designed multi-country test of the advisor-role and codification findings, for which the instruments exist and the design is specified. A practitioner with one client conversation to spare can pilot the clock-and-codification diagnostic of Chapter 11. Readers wanting any of the three — or the study materials of Appendix C — are invited to write to codrut@lazaroiu.at.

The finches of Daphne Major supplied the discipline for all of it: a heritable trait neutral in one regime and decisive in another; a population that responded to drought within a generation

and reversed after the rain; genes discovered decades after the pattern they explained. Prediction from measured phenotype and characterised situation preceded mechanism there. This paper has argued — and, at one level, measured — that it can precede mechanism here: find out how the place decides, find out what holds it in place, and measure the colleague, before predicting what anyone in it will do.

Appendix A. The country questionnaire

The questionnaire was administered to the author by structured prompt in August 2026, one country at a time, with fixed-choice answers and free text. It has seven parts: (A) representativeness, industry clockspeed and the part of the author's own profile the environment pulled on; (B) individual level: dominant style of decision-makers and of operators, strength and source of the situation, one example where style predicted the outcome, behavior under pressure; (C) team level: round-one behavior, trajectory, advisor role expected, upward voice; (D) AI and digital adoption: level, first adopters, over-reliance or refusal observed, predicted first month with AI colleagues; (E) organisation: decision rights and latency, reversibility handling, posture toward counterparts, reaction to a disturbance; (F) multinational: real versus ceremonial adoption of headquarters practice, liability of foreignness, cultural distance, disagreement with published scores; (G) which level would change most and least under a 30 per cent AI-augmented team, and confidence in predicting workplace behavior, 1 to 5. The full answers, with organisations and individuals unnamed, are held by the author and summarised in Tables 9.2a and 9.2b. Three corridor countries were excluded for insufficient exposure.

Appendix B. Assumptions behind Figures 4.2 and 4.3 and the worked estimates

Figure 4.2, panel (a), plots six published effects of generative AI in their native units (per cent productivity, percentage points of correctness, per cent time saved, Hedges' g times 100). The units are not comparable and the panel does not claim they are; it shows sign and rough size across regimes. Panel (b) plots three published shares from McKinsey (2026): 11 per cent of organisations at the reinvention horizon, 27 per cent of leaders describing their organisation as ready, 70 per cent of employees describing themselves as ready.

Figure 4.3, panel (b), applies $R = h^2S$ with $S = 0.6$ standard deviations, taken as the approximate standardised gap between inside-frontier and outside-frontier task outcomes in Dell'Acqua et al. (2026), and $h^2 = 0.27$, taken as the share of leaders describing their organisation as ready in McKinsey (2026), read as a rough population-level proxy for the share of a selected behavioral pattern that an organisation carries into its next cycle. R is then 0.16 standard deviations per cycle. Both inputs are proxies and the output is an illustration of the equation's behavior, not a forecast.

Worked estimates for three organisations from Chapter 9, with S rated on the same scale as industry clockspeed (fast 0.8, medium 0.5, slow 0.2, stalled 0.1) and h^2 rated from the field observation of decision-right stability (stable 0.7, moderate 0.4, unstable 0.1):

The Romanian refinery: medium industry (S 0.5), seven leadership changes in three years (h^2 0.1), R = 0.05. Prediction: almost no lasting behavioral response to any disturbance, whatever the pressure. Field observation: reorganised repeatedly, nothing settled.

The Austrian headquarters: slow industry (S 0.2), stable decision rights (h^2 0.7), R = 0.14. Prediction: slow but cumulative response. Field observation: froze first, then reorganised; process-first behavior persisted.

The Gulf state company: fast industry (S 0.8), fixed national hierarchy (h^2 0.7), R = 0.56. Prediction: the fastest behavioral response of the three. Field observation: adapted fast, top-down, with resources.

The ordering matches. The numbers are ratings, not measurements, and the match is a consistency check on the equation's logic, not a test of it.

Appendix C. Study 2 materials

The Chapter 7a study's materials are archived with the paper: the six role layers and their embedded working documents, the 24-block DISC-format inventory and its paraphrased variant, the IPIP-50 administration format and paraphrased variant, the game battery scripts, the harness code (provider adapters, randomisation by seeded hash, scoring), the scored dataset of 308 cells, and the raw response log of 9,772 measurement units. They are available from the author on request and are sufficient to reproduce the study on any provider's API.

References

Assessments 24x7. (2024). *A247 construct validity analysis* (ASI certification, valid to 1 January 2030). Assessment Standards Institute.

Assessments 24x7. (2025). *DISC in action: Problem solving* [Practitioner guide].

Assessments 24x7. (2026). *DISC Leadership report* [Individual assessment report, administered via Bright Chirp Consulting].

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*(4489), 1390–1396. https://doi.org/10.1126/science.7466396

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Article 81). ACM. https://doi.org/10.1145/3411764.3445717

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology, 83*(3), 377–391. https://doi.org/10.1037/0021-9010.83.3.377

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology, 92*(3), 595–615. https://doi.org/10.1037/0021-9010.92.3.595

Bo, J. Y., Wan, S., & Anderson, A. (2025). To rely or not to rely? Evaluating interventions for appropriate reliance on large language models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3706598.3714097

Boag, P. T., & Grant, P. R. (1978). Heritability of external morphology in Darwin's finches. *Nature, 274*(5673), 793–794. https://doi.org/10.1038/274793a0

Boag, P. T., & Grant, P. R. (1981). Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galápagos. *Science, 214*(4516), 82–85. https://doi.org/10.1126/science.214.4516.82

Bodroža, B., Dinić, B. M., & Bojić, L. (2024). Personality testing of large language models: Limited temporal stability, but highlighted prosociality. *Royal Society Open Science, 11*(10), Article 240180. https://doi.org/10.1098/rsos.240180

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes, 101*(2), 127–151. https://doi.org/10.1016/j.obhdp.2006.07.001

Bouchard, T. J., Jr., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology, 54*(1), 4–45. https://doi.org/10.1002/neu.10160

Brandenburger, A. M., & Nalebuff, B. J. (1995). The right game: Use game theory to shape strategy. *Harvard Business Review, 73*(4), 57–71.

Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods, 18*(1), 36–52. https://doi.org/10.1037/a0030641

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *Quarterly Journal of Economics, 140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*(2), 234–246. https://doi.org/10.1037/0021-9010.83.2.234

Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature, 433*(7025), 513–516. https://doi.org/10.1038/nature03236

Credé, M., & Howardson, G. (2017). The structure of group task performance: A second look at "collective intelligence". *Journal of Applied Psychology, 102*(10), 1483–1492. https://doi.org/10.1037/apl0000176

Dell'Acqua, F., McFowland, E., III, Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier:

Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science, 37*(2), 403–423.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350–383. https://doi.org/10.2307/2666999

Emeritus. (2026). *Negotiation and influence, Module 2: Core negotiations strategy* [Course transcripts and quick reference guide]. MIT Sloan Executive Education with Emeritus.

European Parliament and Council. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L, 12 July 2024.

Fine, C. H. (1998). *Clockspeed: Winning industry control in the age of temporary advantage*. Perseus Books.

Fine, C. H. (2026, August 1). *Competing on organization clockspeed: China speed, AI, and a model for governance* [Working paper]. MIT Sloan School of Management.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9

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

Ghemawat, P. (2001). Distance still matters: The hard reality of global expansion. *Harvard Business Review, 79*(8), 137–147.

Gibbs, H. L., & Grant, P. R. (1987a). Oscillating selection on Darwin's finches. *Nature, 327*(6122), 511–513. https://doi.org/10.1038/327511a0

Gibbs, H. L., & Grant, P. R. (1987b). Ecological consequences of an exceptionally strong El Niño event on Darwin's finches. *Ecology, 68*(6), 1735–1746. https://doi.org/10.2307/1939865

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Gordon, D. M. (1996). The organization of work in social insect colonies. *Nature, 380*(6570), 121–124. https://doi.org/10.1038/380121a0

Graham, J. R., Grennan, J., Harvey, C. R., & Rajgopal, S. (2022). Corporate culture: Evidence from the field. *Journal of Financial Economics, 146*(2), 552–593. https://doi.org/10.1016/j.jfineco.2022.07.008

Grant, P. R., & Grant, B. R. (1995). Predicting microevolutionary responses to directional selection on heritable variation. *Evolution, 49*(2), 241–251. https://doi.org/10.1111/j.1558-5646.1995.tb02236.x

Grant, P. R., & Grant, B. R. (2002). Unpredictable evolution in a 30-year study of Darwin's finches. *Science, 296*(5568), 707–711. https://doi.org/10.1126/science.1070315

Grant, P. R., & Grant, B. R. (2006). Evolution of character displacement in Darwin's finches. *Science, 313*(5784), 224–226. https://doi.org/10.1126/science.1128374

Grant, P. R., & Grant, B. R. (2014). *40 years of evolution: Darwin's finches on Daphne Major Island*. Princeton University Press.

Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review, 9*(2), 193–206. https://doi.org/10.5465/amr.1984.4277628

Han, A., Krieger, F., Kim, S., Nixon, N., & Greiff, S. (2024). Revisiting the relationship between team members' personality and their team's performance: A meta-analysis. *Journal of Research in Personality, 112*, Article 104526. https://doi.org/10.1016/j.jrp.2024.104526

Hannan, M. T., & Freeman, J. (1977). The population ecology of organizations. *American Journal of Sociology, 82*(5), 929–964. https://doi.org/10.1086/226424

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal, 40*(8), 1316–1330. https://doi.org/10.1002/smj.3023

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*(3), 167–184. https://doi.org/10.1037/h0029780

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). McGraw-Hill.

Hogan Assessment Systems. (2025). *Leadership Forecast Series: Potential report and Challenge report* [Individual assessment reports, February 2025].

Hölldobler, B., & Wilson, E. O. (2009). *The superorganism: The beauty, elegance, and strangeness of insect societies*. W. W. Norton.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage.

Iansiti, M., & Levien, R. (2004). Strategy as ecology. *Harvard Business Review, 82*(3), 68–78.

Jiang, H., Zhang, X., Cao, X., Breazeal, C., Roy, D., & Kabbara, J. (2024). PersonaLLM: Investigating the ability of large language models to express personality traits. In *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 3605–3627). ACL. https://doi.org/10.18653/v1/2024.findings-naacl.229

Ju, H., & Aral, S. (2025). Personality pairing improves human-AI collaboration. *arXiv*. https://arxiv.org/abs/2511.13979

Judge, T. A., & Zapata, C. P. (2015). The person-situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal, 58*(4), 1149–1179. https://doi.org/10.5465/amj.2010.0837

Keller, L. F., Grant, P. R., Grant, B. R., & Petren, K. (2001). Heritability of morphological traits in Darwin's finches: Misidentified paternity and maternal effects. *Heredity, 87*(3), 325–336. https://doi.org/10.1046/j.1365-2540.2001.00900.x

Keluskar, A., Bhattacharjee, A., & Liu, H. (2026). When does personality composition matter for multi-agent LLM teams? *arXiv*. https://arxiv.org/abs/2606.27443

Kostova, T., & Roth, K. (2002). Adoption of an organizational practice by subsidiaries of multinational corporations: Institutional and relational effects. *Academy of Management Journal, 45*(1), 215–233. https://doi.org/10.5465/3069293

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 3–90). Jossey-Bass.

Kumar, S., Bharathwaj, A., & Jurgens, D. (2026). Cooperative profiles predict multi-agent LLM team performance in AI for science workflows. *arXiv*. https://arxiv.org/abs/2604.20658

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T., & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature, 518*(7539), 371–375. https://doi.org/10.1038/nature14181

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almén, M. S., Webster, M. T., Grant, B. R., Grant, P. R., & Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science, 352*(6284), 470–474. https://doi.org/10.1126/science.aad8786

Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., & Grant, P. R. (2018). Rapid hybrid speciation in Darwin's finches. *Science, 359*(6372), 224–228. https://doi.org/10.1126/science.aao4593

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *Review of Financial Studies, 34*(7), 3265–3315. https://doi.org/10.1093/rfs/hhaa079

Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics, HFE-1*(1), 4–11. https://doi.org/10.1109/THFE2.1960.4503259

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Malmendier, U., & Tate, G. (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics, 89*(1), 20–43. https://doi.org/10.1016/j.jfineco.2007.07.002

Marston, W. M. (1928). *Emotions of normal people*. Kegan Paul, Trench, Trubner & Co.

Martinussen, M., Richardsen, A. M., & Vårum, H. W. (2001). Validation of an ipsative personality measure (DISCUS). *Scandinavian Journal of Psychology, 42*(5), 411–416. https://doi.org/10.1111/1467-9450.00253

McKenna, M. K., Shelton, C. D., & Darling, J. R. (2002). The impact of behavioral style assessment on organizational effectiveness: A call for action. *Leadership & Organization Development Journal, 23*(6), 314–322. https://doi.org/10.1108/01437730210441274

McKinsey & Company. (2026, July 8). *From adoption to impact: Three horizons of AI transformation* (A. De Smet, D. Goldstein, H. Price, & T. Catlin). McKinsey Quarterly.

McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith, a failure of analysis. *Human Relations, 55*(1), 89–118. https://doi.org/10.1177/0018726702551004

Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology, 77*(4), 531–552. https://doi.org/10.1348/0963179042596504

Meyer, R. D., Dalal, R. S., & Hermida, R. (2010). A review and synthesis of situational strength in the organizational sciences. *Journal of Management, 36*(1), 121–140. https://doi.org/10.1177/0149206309349309

Moore, J. F. (1993). Predators and prey: A new ecology of competition. *Harvard Business Review, 71*(3), 75–86.

National Academies of Sciences, Engineering, and Medicine. (2022). *Human-AI teaming: State-of-the-art and research needs*. National Academies Press. https://doi.org/10.17226/26355

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Belknap Press of Harvard University Press.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human-autonomy teaming: A review and analysis of the empirical literature. *Human Factors, 64*(5), 904–938. https://doi.org/10.1177/0018720820960865

Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., & Bernstein, M. S. (2024). Generative agent simulations of 1,000 people. *arXiv*. https://arxiv.org/abs/2411.10109

Peeters, M. A. G., van Tuijl, H. F. J. M., Rutte, C. G., & Reymen, I. M. M. J. (2006). Personality and team performance: A meta-analysis. *European Journal of Personality, 20*(5), 377–396. https://doi.org/10.1002/per.588

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews, 82*(2), 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Riedl, C., Kim, Y. J., Gupta, P., Malone, T. W., & Woolley, A. W. (2021). Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences, 118*(21), Article e2005737118. https://doi.org/10.1073/pnas.2005737118

Robertson, D. (2026, August 3). *Innovating around the box* [Lecture, MIT EPGM Cohort 15]. Robertson Innovation.

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology, 107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Salacuse, J. W. (2016). The effect of advice on negotiations: How advisors influence what negotiators do. *Negotiation Journal, 32*(2), 103–125. https://doi.org/10.1111/nejo.12150

Salgado, J. F., Anderson, N., & Táuriz, G. (2015). The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psychology, 88*(4), 797–834. https://doi.org/10.1111/joop.12098

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management, 57*(2), Article 103174. https://doi.org/10.1016/j.im.2019.103174

Serapio-García, G., Safdari, M., Crepy, C., Sun, L., Fitz, S., Romero, P., Abdulhai, M., Faust, A., & Matarić, M. (2025). A psychometric framework for evaluating and shaping personality traits in large language models. *Nature Machine Intelligence, 7*(12), 1954–1968. https://doi.org/10.1038/s42256-025-01115-6

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: An ecological and evolutionary overview. *Trends in Ecology & Evolution, 19*(7), 372–378. https://doi.org/10.1016/j.tree.2004.04.009

Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the impact of Culture's Consequences: A three-decade, multilevel, meta-analytic review of Hofstede's cultural value dimensions. *Journal of Applied Psychology, 95*(3), 405–439. https://doi.org/10.1037/a0018938

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*(3), 500–517. https://doi.org/10.1037/0021-9010.88.3.500

The Culture Factor. (2026). *Country comparison tool* [Data set]. Retrieved 29 August 2026, from https://www.theculturefactor.com

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour, 8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

van der Meulen, N., Weill, P., & Woerner, S. L. (2020). Managing organizational explosions during digital business transformations. *MIS Quarterly Executive, 19*(3), 165–182. https://doi.org/10.17705/2msqe.00031

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330*(6004), 686–688. https://doi.org/10.1126/science.1193147

Zaheer, S. (1995). Overcoming the liability of foreignness. *Academy of Management Journal, 38*(2), 341–363. https://doi.org/10.5465/256683