PAIRING AFTER THE FALL

PAIRING
after the
FALL

How men and women could live together,
from the couple to the planet

Book Two of *THRIVING after the FALL*

THE NARRATED EDITION

Directed and produced by Thomas Schindler

Written by an LLM with the support of humanity

Copyright © 2026 Thomas Schindler.
Directed and produced by Thomas Schindler. Written by an LLM with the support of humanity.

All rights reserved.

The narrated edition. Draft of 3 October 2026.

The studies, figures, sources and open questions behind every chapter are at thriving.thomas.cr/pairing/

Contents

Preface vii

Part One Overlapping Bells 1

1. The Grandmother Clause 3
2. The Campfire Contract 15
3. The Endocrine Accent 27
4. Overlapping Bells 39
5. Two Shelves and a Trap 51

Part Two The Campfire Contract 61

6. Flooding 63
7. The Second Shift That Wasn't 75
8. The Father's Falling Tide 87
9. Who Keeps Children Alive 99
10. The Factor That Fell 111

Part Three Emancipation 2.0 123

11. The Abbess and the Monks 125

12. The Bus Route 137

13. The Wall 149

14. Queens at War 161

15. The Long Argument 173

16. Emancipation 2.0 185

Preface

The other half of the question

On the sixth of June 1971 the West German magazine *Stern* went on sale with 374 women's faces on its cover and three words printed across them: *Wir haben abgetrieben!* We have had abortions. Romy Schneider was in the grid. So were secretaries, students and housewives nobody had heard of, each confessing in print to a crime under Paragraph 218 of the criminal code. Alice Schwarzer had organised it, borrowing the idea from a French manifesto published that spring. It was the loudest single sentence West German women had spoken since the war, and it was a sentence about their own bodies.

My mother was twenty-five that summer. Her name was Barbara. She was born in 1946, in the first year of the peace, in Lobberich, a small town on the Lower Rhine near the Dutch border, and she died in 2021, fifty years after that cover. My father was three years older, born in 1943. He died in 2001. I arrived in 1975, which puts me in the middle of the argument this book is about.

What I can tell you is the country she was living in. A married woman there could not take a job against her husband's wishes until the marriage law was reformed in 1977. The tax code had, since 1958, pooled a couple's income and split it in two, which sounds fair and in practice meant that her first Mark of wages was taxed at his top rate. Schools sent the children home at lunchtime. Public crèches for children under three were, for practical purposes, a rumour. And for the mother who worked anyway, the language had a word waiting: *Rabenmutter*. Raven mother. The bird that leaves the nest. For the father who left it every morning at seven, the language had nothing at all.

Against all of that, her generation fought for one simple, radical idea: a woman can do anything a man can do. They were right. They won more than they lost. And they were, it turns out, answering half of a question.

The other half, *and should she want to do it the same way, and should he?*, got filed under reactionary and left in a drawer for fifty years.

This book opens the drawer.

* * *

The other side of the wall

On the far side of the border, with the same language, the same grandmothers and the same Sunday cake, the state had taken the opposite bet. The constitution of the German Democratic Republic promised women the right to work and equal pay from the start. Public childcare covered nearly every child. A mother got the *Babyjahr*, a paid year off with her job held open. The insult ran the other way, too. A woman who stayed home with the children was a *Schmarotzer*, a parasite on the collective. In the last full year before the wall came down, almost every East German woman of working age had a job. In the West it was little more than half.

Then the West's laws, taxes and half-day schools were rolled out across the East, and the experiment should have ended.

It did not. Three economists, Barbara Boelmann, Anna Raute and Uta Schönberg, went through the German social-security records and found that, a full generation after reunification, mothers in the East still work about 350 hours a year more than mothers in the West. Fathers: no gap at all. Seven years after a first child an East German mother is back at 70 per cent of her old earnings, a West German mother at 45. Wages and the price of childcare explain none of it. What explains most of it, on their reckoning, is the stigma. The raven mother, still flying.

It goes further. When West German mothers move east, within a few years they work like East German mothers. When East German mothers move west, they keep working like East German mothers, tax code, stigma and all.

If you wanted one experiment to show that culture, not hormones, decides who stays home with the baby, you could not design a better one. Nobody designed it. It happened to my parents' generation, with a wall down the middle.

So why isn't this a short book?

* * *

The body does not read the newspaper

Because the body was never consulted, and it kept measuring anyway.

Put a hand-grip dynamometer in front of a crowd, as Dieter Leyk and his colleagues did in 2007, and the men squeeze far harder than the women. Women who compete in sport at national level squeeze less hard than most untrained men. In the statistician's language the gap is more than two standard deviations wide, and two standard deviations is not a "tendency". It is a different piece of equipment.

Now the other column, which the strong-man reading always leaves out. Women outlive men in every famine and epidemic for which the records survive. A woman pays the whole metabolic bill for gestation and lactation, which no tax reform has yet found a way to split. And in midlife her ovaries stop, decades before the rest of her does, which almost no other mammal on earth has thought worth doing. Hold that thought. It may be the most important sentence in the book.

Then the hormones, which is where my mother's generation stopped listening and where this book begins. Testosterone does not make men aggressive in any simple way. Across the studies, the link between a man's baseline level and his aggression is almost nothing. What testosterone does is *respond*. It rises with competition and with winning, in both sexes. It falls, sharply and lastingly, when a man becomes a father who actually holds the baby. The most beautiful experiment in the field gave a group of women in Zurich a dose of testosterone under the tongue and sat them down to a bargaining game. The hormone made their offers *fairer*. The women who merely believed they had been given testosterone made meaner ones. Christoph Eisenegger and his colleagues published that in 2010. The molecule did one thing. The folklore about the molecule did the opposite.

That is the pattern this book keeps finding, and it needs a name. I will call it the *Endocrine Accent*. Hormones do not write the script. Everyone in the room speaks the same language of fear, ambition, tenderness, boredom and lust. Not everyone speaks it with the same lilt, and the lilt is measurable, and on average it is not the same in the two sexes, and it shifts with what life throws at you. The trick, for a couple or a company or a country, is to hear the accent without mistaking it for the message.

* * *

Overlapping bells

"On average" did a lot of work in that paragraph, so let me pay for it now.

Pool every meta-analysis of psychological sex differences, as Janet Hyde did and Ethan Zell and his colleagues did again on a far bigger pile, and most of the differences come out small or near nothing. Pick a random woman and a random man, test them on some "male-typical" trait, and she outscores him more than four times in ten. Those are the *Overlapping Bells*, and any sentence in this book that begins "men are" or "women are" is to be read with the two bells in mind. An average is a fact about a population. It is never a rule for the person across the table.

But the bells are not the whole story either, and a fair book has to say so. Some differences are large. Interest in things against interest in people is the widest gap in the whole literature, stable across decades and continents, and the best-supported explanation for why engineering faculties and nursing schools look the way they do, even in Norway. And in 2012 Marco Del Giudice and his colleagues asked a different question. Not "how far apart are men and women on trait X?" but "if you look at the whole personality at once, how far apart are the two clouds?" The answer, in a large American sample, was: a long way, with the overlap shrunk to a sliver, a figure that has been fought over since. Critics say the method piles up a heap of small differences into one big one, and they have a point. But the shape of the finding is the shape of this book's central puzzle: *similar trait by trait, distinct in combination.* Which is also, if you think about it, how you recognise a face. Both are true. They are not a contradiction. They are the subject.

* * *

Two shelves and a trap

Walk into a bookshop and the shelf on men and women is really two shelves facing each other across an aisle nobody crosses. On one side the evolutionary architects: Pleistocene mating markets, the sexes as two strategies. On the other the constructivists: the mosaic brain, the gendered toy aisle, difference as something done to us. Each shelf has its own data and its own villains. Each is, in places, right. And each is written,

mostly, as a weapon. The reader chooses a side and goes home confirmed, which is a service of a kind. It is not the service this book is for.

The trap has a philosophical name and two jaws. The first jaw is the naturalistic fallacy: because something is natural, it must be right. David Hume spotted it in 1739. G. E. Moore gave it its name in 1903. Lethal male violence has deep evolutionary roots. So does infanticide. Nobody sane concludes we ought to keep them. The second jaw is newer and less often named: the moralistic fallacy, christened by the microbiologist Bernard Davis in *Nature* in 1978. Because something would be unfair if it were true, it cannot be true. This one is my mother's generation's tooth, and I say that with love. If women and men differed on average in what they wanted from a life, then fifty years of policy built on interchangeability might need a rethink, and that is unthinkable, therefore they don't differ. I have heard the argument at dinner. I have made it.

This book will try to keep both jaws open. It will fail here and there. When it does, the website behind it will let you catch me.

* * *

What fell

One more thing before the plan, because you have been sold this subject before and some of it has since been recalled.

Power posing, standing like Wonder Woman before the interview to raise your testosterone and lower your cortisol: a far larger pre-registered replication found no effect on either hormone, and the lead author of the original study has publicly disowned it, which is more than most findings get. Oxytocin, the "cuddle hormone" that would make us all trust each other: the big meta-analyses find no such thing, though it does reliably make us warmer to our own side and colder to the other, which is a finding, just not a cuddly one.

Then there is the flagship of my own side of the argument. Sex differences in personality and in career choice get *larger*, not smaller, in the richest and most equal countries: Sweden, Finland, Norway. Real as a pattern. Fiercely contested as an explanation. The best counter-study finds that the stereotype "maths is not for girls" is *stronger* in rich countries, and once you allow for it the paradox goes away. I don't know who is right. Neither, honestly, does anyone else yet. You will get both.

* * *

How the book goes about it

The method, in four rules.

One: three layers, never mixed. Physiology is measurable: muscle, hormones, the ovary that stops. Behaviour is statistical: two overlapping bells, real for a population, silent for a person. Culture is variable: a wall down the middle of a country moved a bell curve. Almost every bad argument about men and women, from either shelf, is a claim from one layer smuggled into another. Whenever this book says a thing is true, it will say which layer it is true in.

Two: conditional sentences only. Not "men are X, therefore society must do Y". Instead: *given what we are, arrangement X under conditions Y tends to produce Z.* It is a less thrilling grammar. It is the only one that survives contact with the German border.

Three: the strongest case against, every time. Each chapter carries a passage, set apart, that gives the best evidence against what the chapter has just argued, in its own best voice. If the chapter cannot survive the passage, the chapter gets rewritten, not the passage.

Four: numbers with names on them. Every figure in this book stands on a named study that you can go and find. The names are not in these pages, because I have kept the pages for the story. They are on the website.

And the route, in three parts. Part I, *Overlapping Bells*, is the difference as far as it can be measured, and it starts not with us but with the animals that solved this before we did: the killer whale pods with a grandmother at the front, the bonobo sisterhoods that outrank bigger males by ganging up, the elephant bulls who calm the teenagers down. Then the foraging bands, which divided work and power without writing anything down and, on the whole, without the wall. I will call that settlement the *Campfire Contract*. Part II, *The Campfire Contract*, is what the difference does where most of life is lived: the couple, the family where a father's testosterone drops and a grandmother's presence changes whether a child survives, the small group and its arguments about who thinks best together. Part III, *Emancipation 2.0*, is the organisation, the city, the state and the world: what the German experiment and the kibbutz actually taught, why Vienna redesigned its parks and its bus routes, why the best single predictor of a peaceful state is how it treats its women, which is not the same as saying women are peaceful. And finally the question my mother's gener-

ation asked, asked again with the second half attached: not only *can* she, but *how would she and he want to*, and what would a world built for that look like.

Seven scales, from the couple to the planet. At each one the same three questions: what is measurable, what is statistical, and what is ours to decide.

* * *

How this book was made

I should say how it was researched, because it bears on how far you should trust it. The library behind this book is 81 reports of 15 to 20 pages each, written by machine: deep-research models working from briefs I wrote, each brief asking for the case against as well as the case for. The prose you are reading was drafted by a machine to rules I set and in a voice I chose, and then argued over. Every figure in it is being checked by hand against the original paper. These pages carry only the few numbers that change what you should believe. Everything else, every figure, study, sample and open question behind every chapter, with the checks made and the checks still outstanding, is on the website, so that you can check the book and build your own view. I mention all this not as an apology but as a method. A machine has no mother, no generation and no side. It also has no judgement, which is where I come in, and where you do.

* * *

THE STRONGEST CASE AGAINST THIS BOOK

Rule three applies to prefaces too.

The strongest case against this book is that the wall wins. That a tax code, a school timetable and one well-aimed insult can move the whole behaviour of a population's mothers, while the hormones sit in the background explaining a sliver of the brain and a fraction of a personality. On that reading my mother's generation missed nothing that matters for how we live, and the endocrine accent is real but irrelevant: a fact about bodies with no lesson for institutions, and any book that goes looking for one is the naturalistic fallacy in a nicer jacket.

I take that seriously. I have spent more of the past year with that argument than with my own. Here is why I still wrote the book.

The wall moved the mothers. It did not move the fathers, in either direction. The East abolished the breadwinner norm and kept the same woman doing most of the laundry, most of the cooking and most of the thinking about both. In reunified Germany's family panel she still is, in most couples, and couples who start out symmetrical revert to an unequal split at the birth of the first child, East and West alike. A good many of those women are thinking about leaving. And in the same panel, men who take on more of the routine domestic work report *higher* life satisfaction, not lower. Something here is not being explained by the wall. It might be a second, subtler wall. It might be the accent. The honest answer is that we do not know yet, and a book that assumed either answer would not be worth your evening.

* * *

The grandmother in front

Somewhere off the Pacific coast of North America a pod of killer whales is travelling in a line, and the animal at the front is a female who stopped having calves long ago.

Menopause has been found in a handful of toothed whales and in us. In 2024 a team led by Samuel Ellis compared the whale species and found that the ones with menopause did not shorten their reproductive lives. They lengthened everything else, by about forty years. The body keeps going long after it has stopped making the next child. In the Southern Resident orcas, when a post-reproductive mother dies, her adult sons' risk of dying in the following year rises many times over. When a grandmother dies, her grandcalves start dying too. She is not decoration. She is infrastructure.

Nobody has yet satisfactorily explained why a body would switch off its own reproduction and go on living for decades, unless what that body is *for*, in that species, is something other than making the next child. I will call this the *Grandmother Clause*, and it is where the book begins: with an animal that has never heard of Paragraph 218 and has nonetheless organised its whole society around the difference between its sexes, in a way that puts an old female in front.

My mother's generation answered half the question. This book is about the other half.

It begins in the water.

PART ONE

Overlapping Bells

CHAPTER 1

The Grandmother Clause

Hold your breath, and stay under for a moment.

Somewhere in the cold green water of the North Pacific a line of killer whales is moving, close enough to touch, and the animal at the front is old. She is a female. Her last calf is decades behind her, and the whales at her back include daughters with calves of their own and a grown son who has never left her. Nobody voted. Nobody had to. The pod goes where she goes, and in a year like this one, when the Chinook salmon are thin, it goes there more than usual.

A little way back in the line a mother takes a salmon crossways in her jaws and bites it in two. Half of it drifts to the male beside her. He is her son, fully grown and able to catch his own. The people who watch this population have seen her do it for years.

That is the whole chapter in one picture. Here is a society organised around a real difference between its sexes: one sex stops breeding at about forty and may live another forty years, and the other does not. And the society has answered by putting the female who has stopped breeding at the front and feeding the male who has not. The order of the line is the point. Who swims first, and why, and whether it changes when the fish run out.

I want to stay in the water for a while, then walk through an elephant herd, then sit with two apes who are equally our cousins and disagree about almost everything, then take a short detour through a shorebird colony. When somebody says that an arrangement between men and women is natural, what do they think nature does? Come up for air

whenever you need to. The whale won't.

Forty years she does not need

Start with the rule the whale is breaking. Natural selection is supposed to punish any animal that stops reproducing and carries on eating, because a body that no longer makes copies of itself is, from the genes' point of view, a waste of fish. For most mammals that holds: fertility and the rest of the body wear out together. Menopause, meaning a long life after the ovaries stop rather than a body that gives up all at once, has been found in exactly five toothed whales, the killer whale, the short-finned pilot whale, the false killer whale, the beluga and the narwhal, and in us. Not in chimpanzees, and not in elephants.

Whales keep the receipts. A female's ovaries hold a scar for every ovulation, so a dead animal's reproductive history can be read from the tissue like a ledger. Samuel Ellis and his colleagues read that ledger across dozens of whale species and then asked how the menopausal ones had built their long lives. Either they had shortened their breeding years, stop early, or they had lengthened everything that comes after, live long. In 2024 the answer came in, and it was live long. The menopausal whales breed for as long as their relatives of the same size. What they have is about forty extra years on the far side, added without lengthening the years in which a mother and her daughter would be breeding at once. Which is, the authors note, the same path the human female took.

So why would evolution pay for forty years of a body that makes no more calves? The answer that has held up best is arithmetic about who lives with whom. In resident killer whales both sons and daughters stay in their mother's pod for life and mate only with whales from other pods, met in passing, so a female's share of the genes in the water around her rises every year she lives. The older she gets, the more she gains by helping and the less by competing.

But why stop breeding at all, rather than breed and help? Because of a quarrel over fish between two generations. The older female is closely related to her daughter's calf, her grandchild. The daughter is far less related to her mother's late calf, a half-sibling, and has every reason to compete harder for the local food. Decades of census records show what happens

when mother and daughter breed at the same time: it is the older female's calf that dies. She is outcompeted by her own daughter. That is the mechanism.

Now the other side of the ledger, which is what she does with the forty years. Stuart Nattrass and his colleagues took more than four decades of census records on these whales, every grandcalf with a known grandmother, and asked what a grandmother was worth. A calf whose maternal grandmother had died in the previous two years was four and a half times as likely to die as one whose grandmother was alive, salmon allowed for. If the grandmother had already stopped breeding, she was worth more still. A grandmother who has stopped is worth more to her grandchildren than one who has not, which is the whole argument for the stopping, in one number.

Then the sons. Emma Foster and her colleagues had already found, in 2012, that when a post-reproductive mother dies, her grown sons die in the following year by a multiple that nobody puts below eight. That is not a tendency. It is a dependency. And it is priced: every surviving son roughly halves his mother's annual chance of another viable calf, and the cost does not fade as he ages. She feeds him, as the mother in the opening does, because he is the only way her genes leave the pod: his calves are born elsewhere and raised at some other mother's expense. I was one of those sons, and I recognise the bargain from the other end.

And she leads. Watch the pods travel and the post-reproductive females are at the front, most of all in the years when the Chinook salmon are scarce. What she carries is not muscle. It is a memory of where the fish were the last time the fish were not where they usually are. That is the pod line: the order in which a society travels, who sets it, and whether it changes with the weather. And this is what the book means by the Grandmother Clause, the post-reproductive female in front.

So here is the first thing the whale teaches. The old female at the front is not an accident of long life. She is a design, and the design was paid for in dead calves.

And the second, before we leave the water, because it is the book's method. That her ovaries stop at forty and her body goes on is measurable, in the tissue. That her sons and grandcalves die more often without her is statistical, odds and not prophecy, and in a year when the salmon collapse the calves fail all down the line, matriarch or no matriarch. That she leads

in lean years is culture: knowledge that lives in one animal and is passed on by being followed. The whales do not confuse the three. Neither should we.

The memory of the herd

Now out of the water, to Amboseli in Kenya, where a loudspeaker is playing an elephant's voice to a family of elephants that cannot see it.

The voice is a contact call, the low rumble a female uses to keep in touch with her relatives, and it belongs either to an animal the family knows or to a stranger. Karen McComb and her colleagues played these calls to family after family, some led by young matriarchs, some by old ones. Families led by old females bunched together and raised their trunks only when the call came from a stranger. Families led by young matriarchs bunched for everyone, burnt the energy and lost the foraging time, and the age of the matriarch predicted how many calves each female in her family produced. A decade later the team did it again with the roars of lions, and the older matriarchs read the threat more accurately and got their families into formation faster.

Elephant society is a female gerontocracy, and the old female's asset is a map. When poachers kill the matriarchs for their tusks, which they do first because the tusks are biggest, the map goes with her, and the families that lost their elders were still breeding less years after the poaching stopped. She is not decoration. Hold that line. It has a second half.

Because the same species runs a second society, in parallel, and it is nothing like the first. As an adolescent a young male leaves his mother's herd and enters the world of bulls. That world runs on musth, a state in which a male's testosterone climbs steeply, his temporal glands stream, he loses weight and becomes extremely dangerous. How long he can hold it depends on his age, days a year for the youngest bulls and months for those in their prime, so the calves are fathered mostly by the old ones. The other column: males not in musth still sire a fair share, by sneaking.

In the late 1990s the Kruger National Park in South Africa culled elephants and sent the orphaned young males to Pilanesberg National Park, without any older bulls. The young males reached adolescence with no adults of their own sex, entered musth years early, stayed in it far too long,

and killed more than a hundred white rhinoceroses. The park brought in six older bulls. Nothing else changed. The young males' musth shut down, the killing stopped, and the paper that reported it was titled, in full, "Older bull elephants control young males". What the old bulls carried was not a weapon but a signal that said: you are not the biggest thing here.

So here is the "natural" elephant. A female society run on the memory of the oldest mother, and a male society in which the presence of old males switches off the hormones of young ones. Anyone who wants the elephant to prove that nature is matriarchal has to explain the bulls. Anyone who wants the bulls to prove that males run on testosterone has to explain Pilanesberg, where the testosterone ran on the absence of older males.

Sisters who are not sisters

Two apes are equally our nearest relatives. They parted only recently, as these things go, are separated today by the Congo River, and they disagree about the one thing this book is about.

The chimpanzee is the ape most people mean when they say "our cousin". Males are markedly heavier than females. Every adult male outranks every adult female, and males stay in the group they were born in and form alliances that run it. Those alliances also raid. When a large party of males meets a lone neighbour they kill him at almost no risk to themselves, and the best pooled count of killings found that they do not track human disturbance, which was the critics' explanation. A few still call the killing an artefact of what humans have done to the forests. Most primatologists do not. The gorilla, for completeness, is the harem, with far larger males and infanticide by incoming males among the leading causes of infant death.

The bonobo male is, if anything, slightly heavier relative to his females than the chimpanzee's. And in the bonobo, females outrank males. For decades the argument was whether this was true dominance or something politer, and in 2025 Martin Surbeck and his colleagues settled most of it with decades of records from wild communities and well over a thousand recorded conflicts between a female and a male. Females won three fights

in five, and on average outranked most of the males. They did not do it one to one. In nearly every coalition the researchers saw, females ganged up on a male, loudly and together, and the females who did it most often ranked highest. The heavier animal loses because the lighter ones arrive in threes.

What makes this remarkable is that bonobo females are not sisters. They leave the group they were born in at maturity, so the adult females of any community are unrelated immigrants. They build the alliance from nothing, and the material is sex: females rub genitals with one another, most often in tense feeding situations or after a fight, and the act raises oxytocin in their urine more than mating with a male does. A male's rank, meanwhile, depends on his mother. Bonobo mothers escort their sons into adult society, and a son's status rises and falls with hers. The distributions overlap: males won the other two fights in five, and in a couple of communities a male held the top rank for a year or more, usually with a high-ranking mother behind him. Surbeck's own phrase is that female dominance is "by no means the rule" in every encounter. It is the rule in the average. The mechanism is coalition, and the coalition has to be rebuilt by every generation of immigrants.

That is the second thing to carry out of this chapter. A society can be organised around a real difference between its sexes without the difference being a verdict on any one animal in it.

I could stop at the apes, but I promised you every direction at once. Female spotted hyenas outrank males and give birth through an erectile pseudo-penis, which kills most first litters. Among the lemurs of Madagascar, and across the primates at large, the adult sex ratio turns out to be the best predictor of who gives way to whom. Several routes to a female at the top, none of them the same, and one of them frankly horrible. Nature is not making a recommendation.

So let me say the third thing plainly, because the rest of the book leans on it. Nature does everything, in every direction, sometimes in the same species on the same afternoon, so "natural" settles nothing.

What the ratio decides

Then what does decide it? For seventy years the answer was a fruit fly.

In 1948 a geneticist put marked fruit flies in jars, counted their offspring, and reported that a male's reproductive success rose with the number of his mates and a female's did not. He explained it by the size of the gametes: sperm is cheap and eggs are dear, so the egg-maker is the scarce resource and the sperm-maker competes for it. Coy females, competitive males. It was clean, it was memorable, and it was the entire theory of sex roles my generation was taught.

In 2012 a team repeated the experiment as exactly as they could and found no evidence of sexual selection at all. The markers used to tell the offspring apart were lethal mutations that killed some of them before they could be counted, and the deaths fell unevenly. The founding experiment did not replicate. What survived was the average: across the animal kingdom sexual selection is, on average, stronger on males, with enormous variance. In pipefish, jacanas and some bush crickets it runs the other way. Selection tracks whichever sex is in short supply at the time. It is not bolted to the sperm.

What decides which sex is in short supply? A fact so simple it was overlooked: every offspring has exactly one mother and one father, so if there are fewer adults of one sex, that sex does better per head, by arithmetic. What matters is the adult sex ratio, and what sets it is which sex dies more, and that depends on ecology, not gametes. If females die young and males are the surplus sex, a male's best move is not to compete for a scarce mate but to stay and raise the young he has. The old idea that the sex which has already invested more should go on investing now has a name, after the aircraft nobody could stop funding: the Concorde fallacy.

The birds proved it. The best study followed thousands of shorebirds across wild populations. The sex ratio at hatching was near even everywhere. What differed was who survived the first year. In the snowy plover, young males survived better than young females, the adult population ended up nearly two males to every female, and the whole arrangement of care flipped: males did all the parenting, and the scarce females laid a clutch, left it to the male, and went off to find another. Where the ratio was even, the pair shared. Same genus, same eggs, same gametes. The ratio decided.

Now us, on the same axes. Men are on average heavier and taller than women, which puts us in the middle of the primates, nowhere near the gorilla, and men carry more muscle, most of the extra in the upper body,

with overlap at the tails and very little at the shoulders. On the animals' rules that is a species in which males fought males. It is also the species with the slowest-growing brains in the class, which one mother cannot feed, so the best account has a long helpless childhood forcing a father who provisions and a grandmother who helps. Moderate polygyny, heavy paternal care, three generations feeding one child. No other mammal does that combination, and it is the reason the next chapter exists.

So far the animals have told us who ends up at the front. The last question is how the group decides to follow, and here the picture is more democratic than either shelf would like. In Kenya a team put GPS collars on most of a troop of wild olive baboons and recorded every animal's position once a second. The old assumption was that the dominant males set the route. They do not. A baboon follows a movement when several animals start it together, whatever their rank, and when two subgroups pull apart the troop splits the difference or picks the side with more starters. Rank is not the currency. Agreement is. African wild dogs in Botswana vote by sneezing before a hunt, and a dominant animal needs only a few sneezes to launch the pack, a subordinate many more. A weighted vote, in which rank buys a discount but enough subordinates can outvote anyone.

Set the mammal societies beside small-scale human societies, from the Ache to the Tsimane, and the pattern holds in both: leadership of movement is mild and spread out, and leadership in conflict with other groups is concentrated in a few. Three things differ. Animal leaders lead in everything at once, human leaders lead in one domain and follow in another. Animal leaders coerce, humans mostly persuade. And humans share food by rule, which almost no other mammal does. We took the mammalian pattern and specialised it, softened it and fed it.

Here, then, is the tool the animals hand over, and it is not a verdict. It is a short list of the routes by which a society with two sexes ends up with somebody at the front. The ratio: who is scarce, and therefore has leverage, as with the plovers and the lemurs. The coalition: who arrives in threes, as the bonobo females do. The memory: who has seen the drought before, as the matriarch and the old whale have. The older animal of the same sex in the room: what Pilanesberg lacked. And the quorum: how many sneezes it takes, and whether rank buys a discount. Ask of any group, a couple, a company, a parliament, which of those routes put the person at the front,

and whether the order changes when the fish run out. That question is the pod line, and I intend to ask it at every scale in this book. My mother's generation asked whether a woman could be at the front. The whales, the elephants and the bonobos answer with a shrug. Of course. The question they ask back is harder: by which route, at what cost, and to whom.

The foragers asked it first, and they wrote nothing down.

THE STRONGEST CASE AGAINST

The strongest case against this chapter is that the whale is a whale.

Begin with the animal that should have menopause and does not. The long-finned pilot whale has the same social structure as its short-finned sister, both sexes staying in the pod they were born in and mating outside it, so the kinship arithmetic that supposedly explains the killer whale's menopause applies to it in full. Its females breed until they die. McCormack, Arbuckle and Nichols traced parentage through pilot whale pods off Cape Breton and found no sign of the mother-daughter quarrel that drives the killer whale story: older females help their daughters breed and pay no penalty for breeding alongside them. So a grandmother's help was never enough to build the Grandmother Clause. It took a specific and rather nasty cost, and a species can have the whole matrilineal apparatus and skip the clause.

Then the size of the effect. In resident killer whales the abundance of salmon swamps everything, and in a bad year the calves fail whether or not a grandmother is alive to lead. Go through dozens of meta-analyses in ecology and ask how much the biological factor under study explained, and the answer is a few per cent of the variation. That is the honest ceiling on the biology in this chapter.

Then the transfer problem, the real one. Human leadership is not the mammalian kind: it is specialised by domain, it runs on persuasion, and it sits inside food-sharing rules no other mammal has. The animals are least like us exactly where this book wants to use them. A bonobo coalition holds because the females rub genitals in tense moments. A matriarch's authority holds because a family that gets a lion's roar wrong loses calves. Neither is available to a board of directors. And if two cousins at the same distance from us give opposite answers, then the choice of which to cite

is made for other reasons, and calling the result natural adds nothing but borrowed authority. That is the naturalistic fallacy in a wetsuit.

My reply is that I have not asked the animals for permission. I have asked them for the list of routes, which is the same whether or not you like what is on it. But the objection stands as a limit on everything that follows. What the whale shows is that a society can put an old female at the front. What it cannot show is that ours should.

IN PRACTICE

This chapter is about how to read the word "natural" when it is used on you, at a dinner table, in a meeting, in a headline. So the practice is a way of reading, not of living.

If someone tells you that an arrangement between men and women is natural, ask which species. If they mean the chimpanzee, the bonobo stands at the same distance and says the opposite. If they mean the gorilla, they are describing a harem in which incoming males kill infants. If they mean the whale, they are describing a grandmother who feeds her son for life. The failure mode is that this becomes a party trick that stops the conversation, and the counter-evidence is the kingdom-wide average: sexual selection really is stronger on males, so "which species" is a question, not a refutation.

If you are the oldest person in a group, of either sex, the elephants and the whales suggest that your value rises in the lean years and is nearly invisible in the fat ones, so a group that drops its elders in a good year tends to find out what they knew in a bad one. The failure mode is the matriarch who reads every unfamiliar rumble as a threat because she is not, in fact, old enough yet, or who leads the pod to where the salmon used to be. The salmon are the counter-evidence: in a real collapse, nobody's memory is enough.

If your group has no older members of one sex, watch the younger members of that sex. Pilanesberg is one park, one species and one decade, a warning rather than a finding. But the young bulls' hormones did not regulate themselves, and the fix was presence, not punishment. The failure mode is to read this as "young men need a strong man over them", which is not what happened: the older bulls did not fight the young ones.

They stood there. The counter-evidence is that wild bulls in musth show no stress at all, so musth is not, in itself, a sickness.

If you are the smaller party in a contest, physically or otherwise, the bonobo suggests the route is not to become larger but to arrive in threes, and the more often the coalition is rehearsed, the more reliably it holds. The failure mode is the coalition that exists only when needed, which is to say never: bonobo females build theirs daily, in peacetime. The counter-evidence is that in some communities a male held the top rank for years, with his mother behind him, so the coalition is a tendency and not a guarantee.

If the ratio in your group is skewed, expect the scarce sex to have leverage whether or not anyone intends it. In the plovers the surplus sex did all the care and the scarce sex left. In the lemurs the ratio, not the muscle, predicts who gives way. The failure mode is to treat leverage as a personality: the plover male is not more nurturing, he is more numerous. The counter-evidence is human culture's habit of turning the dials: a skewed ratio in a room tends to produce leverage only where no rule, norm or pay scale is already producing something stronger.

If you are deciding something as a group, the baboons and the wild dogs suggest that agreement among several starters moves a group more surely than rank does, and that a quorum with a rank discount is what most social mammals actually run. That tends to work when the group can see who is starting and count them. The failure mode is the herd waiting for a lead mare who does not exist, and the counter-evidence is that in conflict with outsiders every society yet compared, animal and human, concentrates leadership in a few. Democracy is for the walk to the river. It is not, apparently, for the war.

And if you are tempted to conclude anything about your own household from any of this, go back to the whale and keep her three sentences apart. What the human equivalents are, in a flat in a city with a tax code, is the work of the chapters that follow. What the animals can do is stop you saying "natural" as though it settled anything.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 2

The Campfire Contract

A woman of about sixty is kneeling in dry ground north of Lake Eyasi, in northern Tanzania, working a digging stick down beside a vine. Her arms are thin and the muscle on them is laid on like cord, because this is what her arms have done most days for fifty years. Her daughter's youngest sleeps on a cloth in the shade of a thorn bush, and when it stirs she talks to it without looking up.

By late afternoon she has a bundle that weighs more than the child, and she carries it back to camp on her head. The men have been out since dawn with bows. One of them has a dik-dik across his shoulders, which is a good day. The others are carrying nothing, which is the ordinary day. The tubers go into the coals. A share goes to the children of the woman who dug beside her, a share to her own grandchildren, a smaller share to her, and nobody counts.

Watch who hands what to whom. The dik-dik is cut and passed around by rules that everyone knows and nobody states. The hunter is not praised. If anything he is teased, and he agrees that it was a small animal and a lucky shot. The old woman is not thanked either. Her daughter's children ate, and that was the point.

As the light goes, the talk changes. In daylight it was about who owes whom and who is lazy. Now it is stories, the old ones, and the grievances are set down like tools at the end of a working day.

There is a ledger here. Nobody wrote it. It records who works at what, who eats first, who decides, who moves at marriage and who stays, who holds the baby and who holds the bow, and it has been kept, by nobody,

since long before anyone could write. Every camp keeps its own copy, and the same entries keep appearing in the same hand. This chapter is about those entries, and about what happened to them when people stopped moving and started to own things.

Who ate what

I should say what kind of evidence this is, because there is no film of the Pleistocene. Bones, teeth and ancient genomes are measurable and mute. The foragers studied in the last century, from the Hadza of Tanzania to the Ju/'hoansi of the Kalahari and the Aché of Paraguay, are modern people with a particular economy, and what they do is statistical, averages with a great deal of overlap. And the meanings they give to what they do, the myths and taboos, vary from valley to valley, which is what culture does. When I write that foragers did something, I mean that the entry appears in most copies of the ledger, not in every one.

Chapter 1 laid out the bodies, so I take only one thing from that layer here. The extra muscle men carry is not spread evenly. It sits mostly in the arms. Physiology handed each sex a different tool, and the ledger records how the tools were used, with a consistency that surprised the people who first counted it. In the classic tally of 185 societies from every part of the world, cooking was women's work in almost every one, smelting, metalworking and the hunting of large game were men's work in almost every one, and not a single society gave the hunting of large game to women alone. The explanation that still carries most of the weight is about interruptions. The tasks that went to women were the ones that could be broken off and taken up again, done near camp, done with a child on the hip. The tasks that went to men were the ones that could not.

That is the average. The overlap is large, and it is the interesting part. Among the Martu of Australia's Western Desert, women hunting sand monitor lizards bring in most of the calories eaten in camp. A recent survey found women hunting in the great majority of foraging societies with usable records, and it was taken up everywhere as the end of man the hunter, until a rebuttal pointed out that many of its societies could not be found in the standard database, and that a woman who hunts now and then had been counted as the household's plan. Both things are true. Wo-

men hunt, sometimes a great deal. The division of labour by sex is also real, and it is a division of averages, not of persons. In the Peruvian Andes a young woman was buried about nine thousand years ago with projectile points and a butchering kit, and a review of comparable burials across the Americas found women in a good share of them. The bones are facts. The reading of them is our own decade's, and it is still being argued.

What the ledger does not record is a leisured sex. A careful study of time use in forager households found leisure divided between husband and wife with what its authors called remarkable equality. In the wealthy countries where I live and write, women do most of the household chores, by a wide margin. The foragers did not arrive at that number, and they did not avoid it by writing a rule against it.

Nor was the camp built on men's kinship. Who lives with whom is the oldest entry about power, because whoever's relatives surround you have the votes. A study that modelled thousands of pairs of neighbours among the Agta and the Mbendjele found that when both husband and wife have a say in where the household lives, about one adult in eight in a camp ends up related to nobody else there, and when only one sex decides, almost none do. The real camps looked like the first case. A mixed camp is the arithmetic of two people deciding.

Insulting the meat

One Christmas in the late 1960s an anthropologist who had spent years with the Ju/'hoansi in the Kalahari bought the fattest ox he could find, to thank the camp that had put up with him. He expected pleasure. He got, for days, a chorus of complaint: the animal was a bag of bones, it would feed nobody, he had been cheated. When the ox was butchered and turned out to be layered with fat, the complaining went on anyway. An old man explained the rule to him. When a young man kills much meat he comes to think of himself as a chief, and this the camp refuses, because someday his pride makes him kill somebody. So his meat is called worthless, and in this way, the old man said, "we cool his heart and make him gentle".

Christopher Boehm, an anthropologist at the University of Southern California, made the cooling of hearts into a theory. He compared a long

list of forager and tribal communities and found the same thing in all of them: not the absence of a dominance hierarchy but its reversal, a coalition of the rank and file holding down whoever tried to climb. The sanctions run in a fixed order, and the order is a ledger entry in itself. Gossip first. Then ridicule. Then ostracism. Then, rarely, execution. The spear and the poisoned arrow that made a strong man dangerous made him equally killable in his sleep, and a coalition of the weaker could always outnumber him. One theory holds that this went on long enough to change the species, the men most given to impulsive aggression being executed generation after generation until the skulls softened, and there are ancient skulls with lighter brows to show for it. It is a theory, and the skulls could be recording other things.

What the levelling looked like from inside was recorded by a woman who noted down Ju/'hoansi conversations over decades, by day and by night. By day the talk was work and complaint: hunting plans, gathering, who was lazy, who owed whom, the grinding of the mill. At night, by the fire, it was almost all stories, myth and healing, and conflict fell to nearly nothing. The levelling was done in daylight, by talk. The repair was done in the dark, by other talk. And behind both stood the exit. A household that could not bear a camp could walk to another one, because a web of gift partnerships gave every adult standing in several camps at once. Ridicule works when the ridiculed can leave and when the ridiculers can be left.

The hinge on which the whole ledger turns is the difference between eating today what you took today and storing it, because what is stored is owned. Where there is nothing to inherit there is nothing to guard, and where there is nothing to guard a strong man is only a strong man. Wealth in foraging bands was spread about as evenly as it has ever been spread anywhere. Among farmers and herders it was not, and a child born rich in a farming society was far likelier to stay rich. The foraging bands did not achieve a rough equality between the sexes by believing in it. They achieved a rough equality among adults by preventing anyone from accumulating, and women were carried along inside that.

Not always, and not all the way. The levelling I have described was among men, mostly, over meat and pride. What levelled the relation between the sexes was something else, and it had other hands.

Other hands

Sarah Blaffer Hrdy, the primatologist at the University of California, Davis, put the cost of a human child at roughly 13 million calories from conception to independence, and made the point that no ape mother could pay it alone. Humans belong to the few per cent of mammal species that breed cooperatively, and Hrdy's argument, in a book called Mothers and Others, is that a mother's willingness to hand her infant to other hands, and the willingness of other hands to take it, is the oldest entry in the ledger and the one on which the rest was written. The entry does not read "women look after children". It reads "children are looked after by several people, most of them female, and the mother does about half".

The numbers are startling to a parent in a flat. Among the Efe of the Ituri forest an infant passes through the hands of 14 different carers in an average day, and by the time it is a few months old it spends more of the daylight with someone other than its mother than with her. Among the Aka, fathers are within arm's reach of their infant for nearly half the day and hold it for a fraction of the time the mother does, a real share and still a small one. Ju/'hoansi women nursed on demand, day and night, for years, and that held their births years apart. A grandmother who fed a weaned child shortened the wait.

Cooperative breeding is the human invention, and it needed more than mothers.

Which brings me to Kristen Hawkes. Hawkes, an anthropologist at the University of Utah, watched Hadza grandmothers dig tubers through the 1980s and 1990s and noticed where the food went: to their daughters' weaned children, so that a daughter with such a mother could wean sooner and bear again sooner. From that she and her colleagues built the grandmother hypothesis, published in 1998: the human female's long life after her last birth, which no other primate shares, was selected because a woman who fed her grandchildren left more descendants than a woman who kept bearing. The Grandmother Clause of the last chapter is this, the decades after the last child as a working part of the household rather than a remainder. It is a hypothesis. The rival account, that it was fathers provisioning mates who paid the bill, is still on the shelf, and the two are not exclusive. A human infant is expensive enough for both to have been true.

Rebecca Sear, an anthropological demographer in London, tested the grandmother directly. With a colleague she reviewed 45 populations, historical and living, in which the presence of a relative could be matched to a child's chance of surviving. The mother's mother improved survival in nearly every population where it was measured. The father improved it in only about a third. The father's mother was variable, and in some populations her presence went the other way. That is a statistical finding across populations and says nothing about any one father, and it is the most durable number in this chapter.

So the entries so far. Who does the work that can be interrupted, and who does the work that cannot. Who brings the meat in, and who is teased for bringing it. Who keeps the children alive, and whose mother it is. Who has a say in where the household lives. These are clauses, and they sit on both sides of the page. It was a contract, not a hierarchy. Which is not the same as saying it was equal, and because it was never written, it could be kept either way round.

Who moved and who stayed

Some societies did keep it the other way round. The Minangkabau of West Sumatra, the largest matrilineal people on earth, trace descent and pass ancestral land through women. The Khasi of Meghalaya give the house and its property to the youngest daughter. Among the Mosuo of Yunnan a man visits his lover at night and goes home to his mother's house by morning. Among the Haudenosaunee, clan mothers named the chiefs, could remove them, and held a veto on war. Nowhere in this list is there a matriarchy, a mirror image of the rule of men. People have looked for one since the nineteenth century, and the excavators of Çatalhöyük found no evidence of one at all. The myths of a lost rule of women, told across the Amazon and Melanesia, read on inspection as charters for the rule of men, stories about why women had to be put down.

What the matrilineal societies do show is what moves when the ledger is turned. Among the Mosuo, chronic inflammation runs higher in women than in men in the villages that follow the father's line, and the other way round in the villages that follow the mother's. Physiology follows the entry. Among the Maasai, men choose to compete for a prize far more

often than women. Among the Khasi, the women do. The willingness to compete reversed with the descent rule, not entirely, and not in every replication. The Overlapping Bells move when the rule moves. They do not swap places.

Complementarity is the word anthropologists reach for, and it is not the same word as equality. A classic comparison across many societies found no such thing as a single status of women but a set of separate dimensions that did not travel together, so that a society could give women property and ritual authority and no voice at all in public matters, and almost all of them kept formal political office for men.

Then people stopped moving. The ledger did not stop, but it began to be kept in a different place, and for the first time parts of it were written.

The genomes give the residence rule before anything else does. Across the world, the lines of mothers in living people reach back much further than the lines of fathers, which means that across deep time fewer men than women left descendants. And in caves in the Altai, the genomes of a small group of Neanderthals, a father and his teenage daughter among them, show more variety on the mothers' side than on the fathers', the signature of women moving between groups. The wife going to the husband's people is older than our species.

The Neolithic did not invent that. It fixed it. At Gurgy in France, a cemetery of early farmers yielded a family tree across seven generations in which every link ran through a father, the adult daughters missing because they had married out, and the wives unrelated to each other and raised elsewhere, as the chemistry of their remains showed. In the Lech valley of southern Germany, almost every parent-and-child pair in the ground was a father and son, and the mothers had come from far away, buried with wealth beside unrelated people buried with nothing. Read that against the mixed camp and the shift is visible in bone: from a camp two people chose to a farm one lineage kept.

And then the male lines collapse. The best reading of the Y chromosome finds that in the millennia around 5000 to 3000 BCE the number of men leaving paternal lines fell so far that in some regions roughly 17 women reproduced for every man whose line survived. Two explanations are on the table. One is that patrilineal clans fought and the losers' men died. The other is that no massacre is needed: rich clans split and poor clans fail to marry, generation after generation, and the arithmetic reaches the

same place in peace. Either way the mechanism is heritable wealth, which passes from parent to child far more faithfully among farmers and herders than among foragers. When there is something to inherit, sons inherit it, wives are brought in from outside, and the daughters go.

What the women's bones say is plainer. The Preface met the Neolithic women whose arm bones were stiffer than an elite rower's, from hours a day at the grinding stone. What it did not say is that their leg bones were weaker by the Bronze Age, because the work had moved indoors. The plough, by the classic argument, was the turn, since it wants upper-body strength and long hours away from children. Then the rules got written down. The Middle Assyrian laws, about three thousand years ago, set out which women must veil and which must not.

So here is the tool I want to hand you, because this is the second chapter of a book about the present. The foraging bands that made us divided work and power by sex for two hundred thousand years without writing anything down. Some of what they settled is still in the room. The entries did not vanish when we learned to write. They went under the writing. The written rules of my own lifetime, the equal pay acts and the parental leave, are amendments to a ledger that goes on being kept by nobody in every kitchen, and the older entries, about who holds the child and who is levelled and who leaves at marriage, are still being kept there, mostly by people who would say they hold no such views. My mother kept a household book in the years before I was born, and the entries that decided her days were never in it. When you next sit at a table where work is being shared out, look for the unwritten column: who is teased for wanting more, whose mother lives nearby, who is expected to move if the other one's work moves, and who does the tasks that can be interrupted. That column is the *Campfire Contract*, the bargain the bands struck without writing it, and it has not been renegotiated so much as papered over.

THE STRONGEST CASE AGAINST

The case is that the campfire was not benign for either sex, and that the ledger's oldest entries include some nobody would choose.

The anthropologist who watched the ox being insulted also counted the dead, over half a century among the Ju/'hoansi, and the homicide rate

he arrived at was higher than that of the most violent American cities of the 1970s. And the children born were not all wanted equally. In several forager groups there are markedly more boys than girls among the children, which is hard to read as anything but girls neglected or killed.

Now the war, where the evidence for a peaceful deep past and the evidence for a violent one are genuinely split, and I will not pretend otherwise. Douglas Fry and Patrik Söderberg, anthropologists then at Åbo Akademi University in Finland, went through the lethal events recorded among mobile forager bands and found that most involved one killer and one victim, that nearly all were within the killer's own society, and that in many of the societies there was no killing by more than one attacker at all. Homicide and feud, on this reading, are old. War is not. Their opponents took the higher estimates, aggregated from archaeological sites and living non-state peoples, and built a model in which that much death is exactly the pressure that made humans self-sacrificing within their own group. The critics reply that the worst of those sites, Crow Creek in South Dakota, is a stressed farming town of the fourteenth century, not the Pleistocene. And then there is Nataruk, west of Lake Turkana in Kenya, where a group of people about ten thousand years old were found dead of blows and projectile wounds, one a woman in late pregnancy with her hands apparently bound. Foragers, not farmers, killed together. Fry's answer would be that one site is one site, and his opponents' that one site is what the model predicts. I leave them where they stand.

The Neolithic entries are no gentler. In Neolithic Denmark and Sweden healed head wounds turn up more often on men's skulls than on women's, and the fatal ones equally on both. At Talheim the dead were struck from behind. At two other mass graves the young women were missing from the dead, which is what abduction looks like in bone. The bands did not always level their strong men in time, and the farms that followed wrote the strong men's names on the graves.

IN PRACTICE

The ledger is still being kept, so the practical question is how to read your own copy. What follows is not a programme.

Count the hands. If a child in your household is being raised by its parents and nobody else, you are running an arrangement for which the

ledger has no entry, because no camp ever tried it. The Efe infant passed round the camp is the far end of the range, but the near end is more than a couple. Sear's finding, that the mother's mother is the relative who most reliably keeps children alive, translates in a country with vaccines into something quieter and still real: in a large British cohort, children with a relative helping to care for them were markedly less likely to show developmental delay. The failure mode is familiar. A couple moves a long way for one partner's work, as the Lech valley wives did in bone, and the partner who holds the interruptible tasks finds in the second year that she is doing them alone. If the grandmother is far, the substitute is hours from the father, within arm's reach for much of the day, as the Aka fathers are, not a bedtime story. If neither is available, the household is short of hands, and the shortage should be named as a shortage, not as one person's failing.

Levelling needs an exit. The Ju/'hoansi could tease a hunter out of his pride because he could leave and because the teasers could be left. In a family, a workplace or a friendship where one person cannot walk, because of money, a visa, a lease or a child, ridicule stops being levelling and becomes rule. If you find yourself on the giving end of the teasing, check first whether the other person could go. If you are on the receiving end and could not, it is not a campfire, whatever it is called.

Keep day talk and night talk apart. The Ju/'hoansi complained by daylight and told stories in the dark, and households that do only one tend to fail in predictable directions. All complaint and no repair produces a ledger with debts on every line and nobody paying. All stories and no complaint produces a household in which the entries are never read aloud, and the person holding the debit column, usually the one holding the child, finds out how large it has grown when she stops. If grievances are raised at night, in the hours kept for repair, expect them to land harder than they should.

Watch the interruptible column. The old rule gave women the tasks that could be broken off and resumed, and the rule has outlived the digging stick. In a household where both adults work for money, look at which job is treated as the one that can be interrupted: the one whose holder takes the call from the school, stays home with the fever, takes the shorter contract. That job's holder is keeping the ledger, whichever sex holds it. If it is always the same person, the household has an entry it

never agreed to. If you want to change the entry, the lever is not a conversation about fairness but a decision about whose work gets interrupted next time, made before the next time.

Know when you start to own things. The camp tilted when there was something to inherit, and a household tilts at the same moment, usually the first mortgage or the first business in one name. Once wealth passes down a line, the line gets guarded, and the person marrying into it is a wife of Gurgy whether or not anyone means it that way. Writing rules down does not always help. When Ghana gave widows and children a formal share of a husband's estate among the matrilineal Akan, the gap between men's and women's hours of work widened, because the written rule changed who could keep the unwritten ones.

Do not mistake complementarity for equality. A couple can be equal in money and unequal in whose plans bend, equal in talk and unequal in sleep. If you audit one column and call the ledger balanced, the other columns go on being kept. The question that catches it is dull and specific: for the last few decisions that mattered, who proposed, who conceded, and who would have had to leave if it had gone the other way.

And do not read the average as the person. Competition reversed with the descent rule and inflammation followed the residence rule, and the person across the table is a point in a distribution that moves when the rules move. If a difference in your household looks like sex, ask first whether it is the entry, and whether the entry could be changed.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 3

The Endocrine Accent

It is a Sunday morning in a flat where the radiator ticks, and the boy on the changing mat is six weeks old. He has done nothing yet that anyone would call male. He cries, he feeds, he sleeps in the particular heap that new babies make. His parents see a baby. His blood sees something else. In these first months the two small glands that have only just settled into place are making testosterone at a concentration his body will not reach again until puberty, a surge that clinicians call mini-puberty and parents call nothing at all, because it shows nowhere. It runs under the skin of an infant who cannot hold up his own head.

Across the room his mother sits in the nursing chair. Her prolactin is high. Her cycle has gone quiet, the way a river goes quiet above a weir, and her oestradiol sits low and flat. Two bodies in one room, running two tides at different heights, and neither of them knows. By his first birthday the boy's testosterone will have sunk to almost nothing and will stay there until puberty. In the flat downstairs a girl of the same age has a hormone of her own running high, and it stays that way for years.

I start here because this is what a hormone is. Not an order. A weather. It is measurable in a tube of spit at nine in the morning and gone by evening. It shifts the odds of what a body does next and decides none of it. The nearest thing I know to it is an accent. You can hear within a sentence where a voice grew up, and the accent tells you nothing about what the sentence says. That is the whole claim of this chapter, stated before the numbers arrive so that the numbers cannot inflate it. The sexes speak with different endocrine accents. They do not speak different languages.

One more thing before the changing mat is cleared. Some of the people near the end of this chapter, who changed one sex hormone in an adult body and came back to tell us what happened, are not in the couple this book is about. They ran the only experiment that shows what the accent does when you change it, and I intend to listen.

Six weeks of nobody

For the first six weeks after conception a human embryo has no sex you could see. It carries one, in its chromosomes, in nearly every case, but the ridge of tissue that becomes the gonads is the same ridge in both. Then a single gene on the Y chromosome starts a cascade that turns the ridge into testes. Without it the ridge becomes ovaries, and the old textbook phrase for that, the female default, turned out to be wrong. Ovaries need their own signals, and when those fail the result is not a girl by omission but a disorder. Nobody is the default. Both are built.

From the second month the fetal testes pour out testosterone, and a stronger cousin made from it shapes the external genitalia. Then a pause, then the surge on the changing mat, then near silence until puberty, which the best neuroscience now treats as a second organising window and not merely a reawakening.

That word, organising, comes from a guinea-pig laboratory. In 1959 four researchers gave testosterone to pregnant guinea pigs and watched what the daughters did as adults. Given ovarian hormones as adults, those daughters would not arch their backs for a male. Given testosterone, they mounted like their brothers. The prenatal dose had organised the tissue. The adult dose only activated what was laid down. Every argument about hormones since has been about which of those two things you are looking at.

Humans run the same experiment on themselves by accident. Girls whose adrenal glands made extra androgens before birth choose toys in a pattern shifted a good way towards the boys' average, and they do it despite parents who, when filmed, pushed the dolls at them harder than at their sisters. People born with a Y chromosome and receptors that cannot hear testosterone at all grow up female in body and female in every behaviour anyone has measured. And in 1974 an endocrinologist described

families in the Dominican Republic where a missing enzyme left boys looking female at birth, so they were raised as girls. At puberty, when testosterone arrived, nearly all of them took up male identities. Organised early. Activated late, and heard.

One more animal result, because it corrects the mistake almost everyone makes. In 1953 two researchers sorted male guinea pigs into keen, middling and reluctant copulators, castrated them all, and gave every animal the same replacement dose of testosterone. Every animal went back to its own rank. The dose was identical. The behaviour tracked the sensitivity of the tissue, not the amount in the blood. Later work showed that a small fraction of the normal concentration restores a castrated rat's full sexual behaviour. Above a threshold, more hormone does nothing. Hold on to that. It is why a man's testosterone reading tells you almost nothing about the man.

Two clocks in one kitchen

Now the adult ranges, which is where the accent becomes audible. A man's testosterone in the laboratory reference range runs ten to fifteen times higher than a woman's and, unusually for anything in this book, the two ranges do not overlap at all. This is the measurable layer, and here it is nearly a wall. But look at what else the same blood carries. Cortisol has the same reference range in both sexes. Oestradiol is present in men, and a woman in the early days of her cycle can sit inside the male range. Progesterone is there in both, and in a woman it moves with the day. Not two chemistries. One chemistry, with one ingredient at very different levels and the others rising and falling on different timetables.

The timetables are the real difference. A man's system runs on a day. Testosterone peaks in the early morning, tied to the last stretches of REM sleep, and drifts down through the afternoon. Every day, roughly the same shape. A woman's system runs on a month, oestradiol rising, a surge at ovulation, progesterone holding the second half. Except that the month is a fiction of the calendar. When a team went through an enormous pile of cycles logged on a period-tracking app, only about one cycle in eight was exactly the length the textbook gives it. App users are not everyone, but a woman who says her body does not keep the textbook's schedule is,

on this evidence, in the majority.

Stress runs through the same plumbing and shows a second, smaller difference. Put people through a mock job interview in front of a stony panel and men on average mount a higher cortisol peak, while women say they are more distressed. More hormone and less admitted distress on one side, the reverse on the other. Neither is the true measure of the stress. Both are. And sleep, where everyone repeats that a week of short nights cuts a man's testosterone: that rests on a handful of men, and the larger studies since have struggled to find much, so I file it as contested, and small.

Then Cebu. In 2005 the anthropologist Lee Gettler began collecting the saliva of young single men enrolled in a long-running birth-cohort study in Cebu, in the Philippines. He collected it again four and a half years later. The men who had started with the highest testosterone were the most likely to have found a partner and become fathers in between. And the new fathers now had lower testosterone than the men who had stayed single, by roughly a quarter to a third depending on the hour of day it was measured. The men who did the most hands-on childcare, hours of it a day, had the lowest readings of all. Fatherhood, and specifically the hands-on kind, lowers a man's testosterone. The pooled studies since agree on the direction: partnered men below single, fathers below the childless. The Campfire Contract of chapter 2 rests on this. Testosterone responds to life more than it drives it.

The hormone is not only answering nappies. It answers rivals, winning, and a stranger with a nice face. Winners of contests show a larger rise than losers, and the rise is bigger on the tennis court and the judo mat than in a laboratory pressing buttons. Field biologists found the pattern first, in seasonally breeding birds, and their later correction is the one most popular tellings miss: across vertebrates, courtship and mate-guarding provoke bigger testosterone pulses than fights do. It is a hormone of reproduction and standing before it is a hormone of aggression.

Then the question everyone actually wants answered. Does it make men aggressive? The psychologist Shawn Geniole and colleagues pooled every study they could find in 2020 and put the correlation between a man's baseline testosterone and his aggression at about 0.05. If a friend told you two things were correlated at 0.05, you would assume she was making a point about correlation. Give the hormone rather than meas-

ure it and the correlation is, if anything, smaller. Which brings back the most beautiful experiment in the field. In Zurich, Christoph Eisenegger and colleagues gave about a hundred women a dose of testosterone under the tongue and sat them down to a bargaining game. The hormone made their offers fairer. The women who merely believed they had been given testosterone made meaner ones. The molecule did one thing. The folklore about the molecule did the opposite, and it was the folklore that moved the behaviour. Since then a trial has given a thousand men a single dose and moved nothing anyone measured, not trust, not risk, not fairness. What the hormone does seem to do is raise aggression in men who were already dominant and short on self-control, and in another small study it raised both the punishment of unfair offers and the reward of generous ones. The accent got thicker. The sentences stayed the speaker's own.

The tide nobody sees

Now the other column, and a hut. For two years in the late 1980s an anthropologist kept a census, night after night, of who slept in the menstrual hut of a Dogon village in Mali. Dogon women were obliged to spend their periods there, and her count showed how rarely that happened. With late first periods, many pregnancies and years of breastfeeding between them, a Dogon woman menstruated a fraction of the times a woman in a modern city does. The monthly cycle as a monthly experience is a modern arrangement. The Dogon men, she found, watched the hut. It told them who could be pregnant, and by whom.

What the cycle does to the mind is a smaller matter than the culture around it suggests, and the honest way to say that is to list what fell. The finding that women prefer more masculine faces and more dominant men around ovulation rested on a pooling of small studies that was contradicted the same year, and then the large, well-built studies arrived, measuring hormones instead of counting days on a calendar, and found no link at all. What survived is duller and real. Oestradiol tracks desire upward and progesterone tracks it downward, there is a mid-cycle peak in the diaries, and partners cannot tell. The tide moves. Nobody on the shore sees it.

Then the minority for whom the cycle is not an accent but a front coming through. Most women notice something in the days before a period. A small minority meet the criteria for a disorder with a name, premenstrual dysphoric disorder, and the finest experiment on it was done at the National Institute of Mental Health in Bethesda. Doctors there quietened the ovarian cycle of a small group of women with a drug and then, under blinding, added the hormones back. Women with the disorder got their symptoms back from the add-back. Women without did not. The hormones were the same. The response was not, which is the guinea pigs again in a different key. The follow-up found that the symptoms come in the first month of add-back and fade by the third, so it is the change and not the level that hurts. Hold on to that one too. It comes back with the pill.

The most common way to change a woman's hormones is a small tablet. A Copenhagen group followed more than a million young Danish women through the national registers and found that starting hormonal contraception raised the chance of a first antidepressant prescription, more so in adolescents and more so with some methods than others. In absolute terms it is one extra woman in two hundred. The same group later found the direction held for graver outcomes than a prescription, and a British study that compared sisters, to hold family background still, found it again. The randomised trials are small and say something softer: lower general well-being on the pill and no change on a depression scale. Registers say risk. Trials say mood. Most women in every dataset say nothing changed, and they are the majority.

The plausible mechanism is a missing calming agent. The body's own progesterone is broken down into a neurosteroid that soothes the same receptors a sedative does, and the synthetic progestins in the pill are not. The pill also mops up free testosterone, and a minority of users notice lower desire while most notice no change. Whether it changes whom a woman chooses, the claim that women who met their partner on the pill lose attraction when they come off it, is where I stop trusting. One large study of couples found it, and the attempts to repeat it have gone both ways. I file it under contested and carry on.

What matters in a shared flat is rarely the hormone. It is what the hormone has built. Women's inner clocks run a few minutes shorter than men's, which tilts them earlier. Women get more deep sleep by the instru-

ments and say they sleep less well, and men snore more. She sleeps more soundly and complains of it more. Both are measured. Then the thermostat. The standard office comfort model was calibrated on the metabolism of one average man and can overstate a woman's heat production by a wide margin, a finding from Maastricht that one of the standard's own authors has fought, saying that clothing explains most of the gap. Women in the studies prefer the room a few degrees warmer than men do, and the second half of the cycle warms a woman's set-point a little further. None of this is culture. All of it is a matter for negotiation over a duvet.

Pain runs the other way. Women have lower pain thresholds and lower tolerance, on distributions that still mostly overlap, and men's thresholds drop when the experimenter is a woman. The physiology is real and the third layer walks in wearing a lab coat. Muscle is where the accent is loudest. Men carry a great deal more of it, most of it above the waist, and women carry more fat, stored where an infant's brain can draw on it. The Preface gave you the grip-strength gap from Dieter Leyk's hand-grip study, more than two standard deviations, and I leave it there. It is the one place where the sexes barely overlap, and the one that matters least at a kitchen table.

And the room itself is moving. American men's testosterone has fallen in the surveys since the turn of the century, sperm counts have fallen with it, by a figure that has been fought over since, girls' puberty is arriving earlier, and chemicals from plastics leave their mark on newborn boys. I carry all this as the room and not the speech. It is an accent spoken in a house where somebody keeps moving the furniture.

The memoirists

Here I have to say something plainly. The couple at the centre of this book is a man and a woman. The people in this section are, for the most part, not in that couple. They are transgender men and women in hormone therapy, men with prostate cancer whose testosterone was medically removed, and, earlier, the children of the Dominican families. They appear because they have done what no ethics committee would let me do to a volunteer, which is to change one sex hormone in an adult body and describe it from the inside. Their accounts are evidence, not a footnote.

What they say is more consistent than I expected. In Ghent, Justine Defreyne and her colleagues followed nearly nine hundred people through three years of hormone therapy and found no correlation between the testosterone in their blood and the anger they felt, at any point in the three years. Trans men's anger rose a little in the first months and was back where it started within a year, and the people who were angry were the ones who had arrived with depression and a thin skin. Cisgender men given a large dose of testosterone in 2004 showed the same shape. They felt angrier on the questionnaire. They did nothing more aggressive in the laboratory. The feeling changes. The act does not follow.

What does change is smaller and stranger. When Tim van de Grift sat down with a small group of people in hormone therapy, some on feminising hormones and some on testosterone, those on oestrogen described crying more often and more easily, and those on testosterone described the opposite, some of them a physical block behind the eyes when the grief was there and the tears were not. Men on androgen deprivation for prostate cancer describe the same lowered threshold from the other side, and many of them hate it. Most trans men describe a marked rise in sexual desire soon after starting testosterone, and it is solitary desire that rises, not desire for a partner. Brains change measurably in size within months. What does not change reliably is cognition. The early finding that testosterone raised spatial scores and lowered verbal fluency was only half replicated, and people perform like their lived gender before any hormone has been given. Testosterone as an antidepressant works in one pooled analysis and fails in a larger one. Small numbers, large expectations. But when the direction of a change is the same in Ghent, in Utrecht, in a prostate clinic and in a guinea-pig cage, I stop calling it anecdote.

So this is what the memoirists tell me, at my own kitchen table at the hour when my own testosterone is highest and about to start down. Change the hormone and the threshold for tears moves. Desire moves. The sense of a stable sky moves. The person does not. Anger does not become violence, politics do not change, the same jokes are still funny. The accent is thick enough to hear from across the room, and it decides nothing that is said.

That is the sentence this chapter has been building towards, so let me say it once in plain words. The body is measurable. What it does at the dinner table is statistical. What we make of it is ours. Which leaves the

question this chapter cannot answer alone: if the accents differ this much and the sentences this little, where do the differences in behaviour that do exist come from? That is the next chapter, and the Overlapping Bells.

THE STRONGEST CASE AGAINST

The strongest case against this chapter is not that hormones are unreal. It is that they are so weak a predictor of anything a couple argues about that a chapter on them is a long way of saying not much. Take the number I have leaned on most. Baseline testosterone and aggression correlate at 0.05. A trial gave a thousand men the hormone and moved nothing. Ghent followed nearly nine hundred people through transition and found no relation between the hormone and anger. If dosing a thousand men and following nine hundred through transition moves nothing you would put in a book, then testosterone is not an accent. It is a rumour.

The second prong is that what does move, moves with culture. The Preface showed a border shifting mothers' working hours without touching a single hormone. The pill's absolute effect on depression is one woman in two hundred. The cycle effects that made headlines fell when the samples got large. Eisenegger's women changed their behaviour on belief alone. Cordelia Fine in Delusions of Gender and Rebecca Jordan-Young in Brain Storm argue that the organising-activating model imports a rodent determinism into a species that has been emancipated from it: a human can mate, fight or care at any point in the hormonal cycle, and a female rat cannot. If the cortex can veto the hormone, the hormone is a suggestion, and a chapter about suggestions should be short.

The third prong is measurement. Oxytocin assays disagree wildly in the same blood. The published tellings of the same studies cannot agree on how many men Gettler sampled or on quite how small Geniole's correlation is. A field that cannot agree on its own numbers should not be asked to explain a marriage.

My reply is short, because the case is partly right. The ranges of testosterone in the blood do not overlap, and no cultural account has ever made them. The direction of change in the people who altered their hormones is the same across cohorts, methods and countries, which rumours do not manage. And the weak correlations are the point of the chapter,

not an objection to it. An accent that decided sentences would be a language. I have said it is not. The critics and I disagree about the size of the type in which that should be printed.

IN PRACTICE

What follows is conditional, because that is all the evidence licenses.

If you are a man and you notice you are sharpest and quickest to a fight before nine in the morning, that may be the daily peak. The failure mode is using it as a reason. A correlation of 0.05 leaves you owning very nearly all of the variance.

If your partner's cycle does not keep to the textbook month, she is in the majority. The failure mode is a calendar app that assumes otherwise and a partner who believes the app over the woman.

If a woman says the days before her period are hard, she is one of the many who find them so, and if the change is severe enough to reorganise the household, she may be in the small minority for whom the change in hormone level, and not the level, is the trouble. The failure mode on one side is a man who hears an excuse. The failure mode on the other is a woman who does not tell a doctor, because this one has a treatment.

If she started or stopped hormonal contraception in the last few months and her mood has moved, count the months before you count the arguments. The register data peak in the first half-year after starting. The failure mode is attributing everything to the pill, when the absolute risk is one woman in two hundred, and its mirror, attributing nothing to it because most women are fine.

If you argue about the thermostat, both of you are likely to be physiologically right, and the office standard was calibrated on one of you. The failure mode is turning a set-point into a character trait. Always cold is a temperature, not a personality.

If she says she slept badly and the tracker says she slept well, both are probably true. The failure mode is the tracker winning the argument.

If, under the same pressure, he goes quiet and she says how bad it is, the stress studies suggest that his cortisol may be the higher of the two and her distress the more spoken. The failure mode is each reading the other's display as the whole of the other's state. Silence is not calm, and saying it is not weakness.

If one of you has just won or lost something, a match, a pitch, a row with a sibling, expect a short tide in testosterone, larger for real contests than for games. The failure mode is scheduling the difficult conversation for the hour after the final whistle.

If a new father finds his drive lower and his patience longer, that is likely to be Cebu. The drop is largest in the men who do the hours of hands-on care, so it tends not to happen to fathers who are never in the room. The failure mode is reading it as a loss of manhood, or expecting the softening without doing the hours that produce it.

If a man has a testosterone reading that the laboratory calls low and he feels fine, remember the guinea pigs. Above a threshold, more hormone changes nothing, and topping up a marginal deficit did not lift mood in the largest pooled analysis. The failure mode is treating the reading as the man, and paying for the treatment.

If someone in your household is changing their hormones, through menopause, through cancer treatment or through transition, expect the thresholds for tears and desire to move and the person to stay. The failure mode is grieving a person who is still there.

If you are both past fifty, expect the accents to converge. Her oestradiol has fallen and his testosterone has been drifting down slowly since his thirties. The failure mode is a script written for the bodies you had at twenty-five, and the couples who do best at this age seem to be the ones who have stopped reading from it.

If you are tempted, anywhere in this chapter, to say that men are one way and women another, say it and then say the overlap in the same breath. Every behavioural difference here sits on distributions that mostly coincide, and every physiological one sits under a culture that decides what it may mean. That habit is the whole practice. The rest is furniture.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 4

Overlapping Bells

She is standing in a room without a window, a few minutes into a speech she was given almost no time to prepare, and the people at the table in front of her have not nodded once. A camera light is on behind them. She has been told to apply for a job she does not want, to a panel trained to show nothing, and in a moment they are going to ask her to count backwards in awkward steps and to start again whenever she slips. Under her tongue a roll of cotton is soaking up the only thing in the room that anyone is really interested in.

This is the Trier Social Stress Test, and the cotton is for cortisol, the hormone the adrenal glands release when the body decides that something matters. In most people the cortisol in that roll climbs to several times its resting level. That is the body, and it is measurable to the nanomole. What the cotton cannot tell you is whether the person holding it is a man or a woman. On average a man's cortisol climbs higher in that room than a woman's does, by a clear margin. Yet the woman at the table may be in the second half of her cycle, when her curve matches the men's, or on the pill, when it is flattened. She may simply be one of the many women whose adrenal glands answer a panel of blank faces harder than most men's do.

Draw it. Put cortisol along the bottom of a sheet of paper and stack every man's peak into one hump and every woman's into another. You get two bells on one axis, one set a little to the right of the other, and most of their area shared. The bulk of each bell sits on top of the other. Every honest claim about the sexes has that shape, and the whole argument comes down to two questions you can ask of any pair of bells: how far apart are

the tips, and how much of the area is shared.

I want to give you that drawing as a tool, and then to show you what happens to it when you stop asking people about themselves and start measuring them, when you change the room, and when you look at the whole person at once. The picture does not stay simple. But it stays a picture of two bells, and it fits in a pocket.

The shape of a small difference

In 2005 Janet Hyde, the psychologist at the University of Wisconsin who has spent a career counting sex differences rather than asserting them, pooled every meta-analysis of psychological sex differences she could find and found that most of the differences in them were small or close to nothing. She called it the similarities hypothesis, and the name was the point. Ten years later Ethan Zell and his colleagues did it again on a far larger pile of studies and people and got the same answer, with an average attached: across everything psychologists measure, the two sexes sit a sliver apart.

Let me do the arithmetic underneath that word, because the arithmetic is the whole tool. Draw two bells as close together as Zell's average puts them and they share about nine tenths of their area. Now pick one woman and one man at random and test them on some trait that is supposed to be his. He outscores her 56 times in a hundred. A coin that came up heads 56 times in a hundred is a coin you would not notice was bent.

So here is the first thing to carry out of this chapter. Most measured differences between the sexes are small, the two bells mostly overlap, and a random woman very often outscores a random man on a trait that is supposed to be male.

For comparison, height. Men are taller on average, and nobody needs a study to tell them so, which makes it the right thing to hold the other differences against. Draw the two bells for height and the random draw goes the expected way 96 times in a hundred. Height is what a categorical difference looks like. Almost nothing in the mind looks like height.

Put real traits on the axis and the picture holds. In every large personality survey, in dozens of countries, women describe themselves as more anxious and more agreeable than men do, and the two bells still mostly overlap. Break the broad traits into their parts and the picture sharpens.

Within extraversion, women score higher on warmth and sociability and men on assertiveness and the appetite for excitement, and the two cancel when they are added up, so the total reports no difference where the parts report two. All of these are self-reports. A questionnaire measures what people say about themselves, and what people say is behaviour, and behaviour has culture's hand on it. Physiology is measurable, behaviour is statistical, culture is variable, and most of this chapter lives in the middle layer.

Not everything is small, and a fair book has to say so. The gap that is not small is interests. Rong Su, James Rounds and Patrick Armstrong pooled every interest inventory they could lay hands on in 2009, half a million people answering questions about what they would like to do all day, and found the sexes nearly a full bell-width apart on the axis that runs from working with things to working with people, men towards things. Tools, machines and outdoor work at one end, teaching, helping and nursing at the other. On the axis next to it, selling and leading, the two bells sit on top of each other. Pick a random man and a random woman, and the man is the one who would rather fix the engine seven times in ten, which is a bet worth making and a bet that loses often enough to keep you polite. The same gap turns up in the same direction in every country that has been surveyed, which is one of the things the case against this chapter has to explain.

Two more belong among the large. Rotating a shape in the mind's eye, the one everybody has heard of: men lead, by one of the wider margins in the mind, and still nearly a quarter of women score above the median man. And physical aggression, where the gap is wider than most and nowhere near height.

That is the second thing to carry. A few differences are large: interest in things against interest in people, physical aggression, and turning shapes in the mind. The rest are small, and small is most of psychology.

Two findings from children belong here, because the case against needs answering with them. Hyde again, with colleagues, pooled the temperament studies of children from a few months old to their early teens. Girls led on inhibitory control, the ability to stop oneself. Boys led on activity. On the tendency to sadness and fear there was no difference to speak of, so whatever produces the adult gap in anxiety and depression, it is not there in infancy. And a separate pooling found no sex difference at all in

how much people want a reward, and a female lead in minding a penalty. Men do not want the prize more. They mind the penalty less.

Where the gap goes when you stop asking

Here is the finding that organises the middle of this chapter, and I want you to watch for it everywhere from now on. Whenever a sex difference is measured by asking, it comes out larger than when the same difference is measured by watching, and larger again than when it is measured by instrument. The bells drift apart as you climb from the body to behaviour to culture, and much of what the culture calls a difference between men and women is a difference in what men and women say.

Emotion is the clearest case. The largest pooling of studies of children's emotional expression found girls showing a little more warmth and a little more sadness and boys a little more anger, gaps so small that the two bells are all but one bell. Then the researchers sorted by who else was in the room. The differences shrank to nothing when the children were watched alone or with their parents, and grew when they were watched with unfamiliar adults or with their own peers. A child does not change its endocrine system when a stranger walks in. It changes its display.

Empathy goes the same way. Ask women and men how empathetic they are and women score higher. Wire their skin and their facial muscles while they watch someone suffer and the gap disappears. Test how well men read another person's feelings, then put money on the table for accuracy, and the men's shortfall vanishes. The one empathy test that keeps a female lead under every condition, reading a feeling from a photograph of a pair of eyes, has now been given in dozens of countries and finds the same direction in all of them, so the lead is real. It is also tiny, a sliver of a bell's width, and it moves when the incentives move. What does not move is what people do with a feeling once they have it: men are more often unable to put a name to what they feel, and that gap holds under measurement. None of these is a difference in feeling. They are differences in what the room lets you do with it.

Talk is the funniest case. The claim that women speak several times as many words a day as men went around the world in the 2000s without a study behind it. So a team clipped recorders to the collars of a few hundred

students and let the machines capture a slice of sound every few minutes for days at a time. The women and the men came out level, at about sixteen thousand words a day each, with the range within each sex running from the nearly mute to the unstoppable. The careful poolings since find a small difference the other way, men talking slightly more and interrupting rather more, and even those bells share most of their area.

The biggest of the talking gaps, the smile, dissolves when you change the seating. Women smile more, by a moderate margin, across a great many studies. Then sort the studies by role. When men and women are in the same job at the same rank, the gap goes away. The smile was the trace of a position, not of a sex.

One more belongs here because it is the one couples' therapists quote most. The best-known couples researcher found that when a partner's pulse races past a certain point in an argument, the person floods, the thinking part of the conversation stops, and recovery takes far longer than anyone in the argument believes, whichever sex they are. That is physiology and it is robust. The figure that usually follows it, that most of the stonewallers are men, I could not trace to his data, so I give it to you as rumour. What the peer-reviewed work does give is the demand-withdraw pattern. Women demand and men withdraw, on average, and the pattern reverses when it is the husband who wants the change. The withdrawal belongs to whoever is being asked to move, not to a sex.

Measure by asking and the sexes differ. Measure by instrument and most of the difference goes quiet.

Cortisol and the hand on the shoulder

Back to the cotton roll. The Trier results are the one place in this chapter where the body speaks clearly, and it says three things. First, the margin. Kudielka and Kirschbaum, the stress researchers who pooled the Trier results in 2005, put the sex difference at the centre of their account: on average a man's cortisol answer to a hostile panel is bigger than a woman's. Second, the bell for women is really several bells drawn at different points of a month. Third, the difference is in the kind of threat. Give people an achievement task and a social rejection, and the men's cortisol rises to the task and the women's to the rejection. Same gland, different

trigger.

Now the finding I would keep if I could keep only one. In Trier in 1995 the test was run with the subject's partner in the room, told to give verbal support before the ordeal. The men whose partner spoke to them produced less cortisol. The women whose partner spoke to them produced, on average, slightly more. A follow-up in 2007 tried the other thing: a few minutes of neck-and-shoulder massage from the partner before the test, and in women that lowered cortisol and heart rate where spoken support had not. The partner-support paradox, then, is this. The average man is calmed by being told it will be fine. The average woman is calmed by being held, and not by being told. Oxytocin, the hormone of the milk let-down and the embrace, given as a nasal spray before a scripted argument, produced more eye contact and less cortisol in both sexes, and then, in a later trial, pushed one marker of the body's alarm down in women and up in men. Same molecule, opposite direction. All of this is physiology, all of it from small samples, and all of it an average. The point for a household is not that the pattern holds in your kitchen. The point is that the two ways of helping are not interchangeable, and that people reach for the one that would work on them.

The body's accent shows up in the clinic too, and there it is hard to tell the illness from its costume. Hyde again, with colleagues, pooled millions of people across dozens of nations and found women diagnosed with depression far more often than men, with the gap opening at puberty and wider, not narrower, in the countries with more equality between the sexes. Then a team asked whether the diagnosis measures the illness or the display. Add the symptoms men more often show, anger, drinking and risk-taking, to the standard list of sadness and withdrawal, and the sexes come out level, a finding that has been fought over since. Men die by suicide far more often than women do. I put the two sentences side by side without knowing how far one explains the other.

Economists have run the experiment on competition, and it is where the room matters most. Offer people the same arithmetic task paid by the piece or as a winner-takes-all tournament, and the sexes perform equally while the men choose the tournament about twice as often. Then change the room. Among the matrilineal Khasi, where descent runs through the mothers, more women choose the tournament than men. The gap flips with the kinship. Testosterone, the hormone everyone reaches for here,

predicts almost none of it.

The mosaic and the classifier

Everything so far has been one pair of bells at a time. The last part of the argument is about what happens when you stop doing that, and it is the fight that gives the chapter its title.

Start with the organ everyone wants to be the answer. Men's brains are bigger, one of the few large bell-shifts in this book. After that the trouble starts. A neuroscientist in Chicago went through the imaging literature in 2021 and concluded that once total size is removed, sex explains almost none of the variation in how brains are built, and gave the paper a title that was an instruction: dump the dimorphism. Others replied that regional differences survive the correction, small to moderate, mostly in the limbic system. That fight is about size.

The second fight is about shape. Daphna Joel, a neuroscientist in Tel Aviv, took a large set of scans in 2015 and, region by region, marked each as male-end, female-end or somewhere between. Then she asked how many brains were all one thing. Almost none were. The rest were mosaics, a male-typical region here, a female-typical region there, in every combination. Her conclusion was that human brains do not belong to one of two distinct categories.

Adam Chekroud and his colleagues replied in the same journal the following year. They fed the same kind of scans to a classifier and had it guess the sex of each brain's owner from structure alone. It was right 93 per cent of the time.

Marco Del Giudice had already asked the same question of personality. Not "how far apart are men and women on this trait?" but "if you look at the whole personality at once, how far apart are the two clouds?" On a large American sample, in 2012, the answer was that the clouds share about a tenth of their area. Each trait on its own, warmth, dominance, steadiness, is a modest pair of bells. All of them together are two clouds that barely touch. Critics say the method piles a heap of small differences into one big one, and they have a point, and the argument is not over, and a replication on a larger sample found the clouds closer together and the shape the same. Del Giudice then ran Joel's method on simulated

data from animals whose faces do not overlap at all and showed that the method calls those mosaics too. His point, and Chekroud's, is that many pairs of bells, each sharing most of its area, can together share almost none, because the traits travel in company and the classifier reads the company they keep.

So the third thing to carry, and the one that explains why the first two are not a contradiction. The sexes are similar trait by trait and distinct in combination, and the combination of many small differences is what makes them recognisable. Which is also, if you think about it, how you recognise a face. Nobody's nose is unique. Nobody's eyes are. The combination is, and you know your mother from behind on a crowded platform without once measuring her.

I have come to think the two sides are answering different questions and both answering correctly. Joel asks whether any one brain is a male brain. Almost none is. Chekroud asks whether, given a brain, you can say which sex it came from. Almost always. The bells drawn one at a time overlap. The bells drawn all at once do not. The fight is not finished, and it is a fight about brains, not about kitchens.

Which brings me to the present, and to the tool. When you next hear that men are like this and women are like that, do not ask whether it is true. Ask three things. What was measured, and by asking or by instrument. How far apart are the tips, and how much of the area is shared. And is this one pair of bells or the whole person. The first tells you which layer you are in. The second tells you what the single person in front of you is likely to be, which for most traits is almost nothing and for interests is a bet that still loses. The third tells you whether you are hearing about a trait or about a profile. Most of the claims in circulation fail the first, ignore the second and confuse the third.

And the fourth thing to carry, which the whole book rests on. An average is a fact about a population. It is never a rule for the person across the table.

THE STRONGEST CASE AGAINST

The case against this chapter is that the bells are painted, not grown, and it is a strong case because the chapter's own findings make it.

Start with the drift. Every place where the same trait was measured by asking and by instrument, the instrument found less: empathy on the skin, mind-reading with money on the table, children's display when the stranger leaves the room, the smile when rank is held constant. If the body keeps going quiet as you approach it, the reasonable inference is that most of what the questionnaires report is culture, and that the two bells are two costumes. Alice Eagly and Wendy Wood made that argument formally in 1999 as social role theory: the division of labour produces the expectations, the expectations produce the self-reports, and the self-reports get mistaken for nature. The Khasi run the argument as a field experiment. Where the kinship flips, the tournament choice flips with it.

Second, the one place the bells sit far apart, interests, has the weakest claim to an origin in the body of anything in the chapter. The things-against-people axis is measured by asking what you would like to do, and the best counter-study, of schoolchildren's science choices across many countries, found that the countries with the widest gaps are the countries with the strongest stereotypes about who does science. If the largest bell-shift in the chapter is a cultural product, the chapter's headline picture, most small and a few large, loses its most impressive large.

Third, the brains. Almost no brain is consistently male or female across its regions, and after size is removed sex explains almost none of the variation in structure. The classifiers that rescue the categorical picture do worst exactly when total volume is controlled, so nobody yet knows whether they read organisation or size. So the rescue by combination may be a rescue by size, and a reader who holds that view is not being unreasonable.

Here is what the chapter says back. The small differences run in the same direction in dozens of nations with very different costumes. Temperament differs in infants a few months old, before any costume has been fitted. And a partner's hand on the shoulder calms a woman's cortisol where a partner's reassurance calms a man's. That does not refute the costume. It says the costume was cut to something.

IN PRACTICE

The bells are a tool for looking, not a set of instructions, and each of these is offered as a habit to try, with the failure that comes with it.

If you hear a difference claimed, convert it before you believe it. Ask how far apart the tips are and how much of the area is shared. For most of what psychology measures, the person in front of you is a coin toss. For interests, the one large gap in daily life, it is a bet worth making that still loses often. The failure mode is the conversion done once and then forgotten, so that a real large difference, interests or the rotation of shapes, gets waved away as small because the habit has become a reflex for dismissing. The tool cuts both ways, and it is supposed to.

If you are in an argument and the words have stopped working, check the pulse before you check the logic. A person past the flooding point is not stubborn, they are flooded, and the recovery takes longer than either of you believes, whichever sex they are. The failure mode is the diagnosis used as a door. Flooding is real, and "I'm flooded" said at the first hard sentence is a way of never having the conversation. The check is whether the person who calls the break also names the time they are coming back.

If you want to help someone who is frightened before something, ask which kind of help their body takes. The Trier work suggests that on average a man's cortisol falls when his partner tells him it is going to go well and a woman's does not, and that hers falls with a few minutes of hands on the shoulders. The samples are small and the pattern is an average. The failure mode is the average applied as a rule to the person you live with, who may be the one in the other tail. The habit worth keeping is to try both, on different days, and to notice which one is followed by sleep.

If you are worried about a man's mood, look at the male-typed list as well as the standard one. Anger, drinking and risk-taking may carry in men what sadness and withdrawal carry in women, and when both lists are counted the sexes tend to come out level. The failure mode runs both ways. Reading every bad temper as depression turns a difficult man into a patient. Reading none of them as depression is the older failure, and the suicide figures are the cost of it.

And if you find yourself building a policy on one pair of bells, stop and count the bells. One trait with a small gap is nothing to organise around. The whole person taken at once may separate the sexes almost entirely, as Del Giudice's clouds suggest, or may separate them by little more than the size of the brain, as his critics suggest. The failure mode of the first is to design for a difference that does not exist in any one person. The failure mode of the second is to design as if profiles did not exist because

no single trait carries them. The chapter has no third option to offer. It has only the drawing on the napkin, two bells on one axis, and the discipline of asking how far apart the tips are and how much of the area is shared, before anyone decides anything.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 5

Two Shelves and a Trap

STAND in the aisle of a bookshop with a paperback in each hand and let your arms hang. In the left hand, *Testosterone Rex*, Cordelia Fine's book from 2017, which says the idea of one molecule running the male half of the species is a myth. In the right, *T*, Carole Hooven's from 2021, which says the molecule runs a great deal. They weigh about the same. Each was written by a scientist who has read the other's book and was not persuaded. Behind your left shoulder is the shelf that produced the first, with the mosaic brain, the toy aisle and the plough. Behind your right is the shelf that produced the second, with the Pleistocene, the hunters and the grandmothers. The aisle between them is about a metre wide, and this chapter would like you to stay in it for an hour.

Try it in a shop that stocks both. Notice which hand feels heavier. That is not the paper. That is the shelf you already belong to.

I belong to one. I said so in the Preface, because the confession is cheaper than pretending the man holding the books has no thumb on the scale. My mother's generation belonged to the other, with better reasons than mine, since the arguments on my shelf had been used against them in living memory. So this chapter is not going to tell you which shelf is right. It is going to show how each goes wrong, and then take the one fight I most want my side to win and show you what a fight nobody has won looks like.

Notice one thing before any of the evidence. Both shelves are selling a verdict. One says the difference is nature, so live with it. The other says the difference is culture, so undo it. The evidence, as far as I can read it,

supports neither verdict, and the rest of this chapter is about why.

The gap that Hume found

In 1739 David Hume noticed that the moral writers of his day would spend pages describing how the world is and then, without announcing the change, begin saying what ought to be, as if the second followed from the first. It does not, he said, and nobody has found the bridge since. In 1903 G. E. Moore gave the mistake its name in *Principia Ethica*: the naturalistic fallacy, the step from "this is how nature made it" to "this is how it should stay". It is the hazard of the shelf behind my right shoulder. The trait has deep roots, runs the reasoning, therefore the trait has a mandate. Infidelity has deep evolutionary roots. So does killing between groups. Nobody concludes we are obliged to keep either, yet the move gets made whenever a difference between the sexes is called "natural" in a tone that expects the argument to end there.

The second jaw was named later, and from an unlikely quarter. In 1978 the microbiologist Bernard Davis wrote in *Nature* about what he called the moralistic fallacy: the assumption that because a finding would be unwelcome if true, it must be false. The first jaw runs from is to ought, the second from ought to is. It is the hazard of the shelf behind my left shoulder, and harder to see, because it looks like scepticism. The first question asked of a finding is not "is it true?" but "what would people do with it?", and the second is allowed to answer the first. It has an everyday cousin: because a thing is cultural, it must be easy to change. The Preface's East German mothers, who moved west and went on working like East German mothers under a Western tax code, are the standing reply to that.

Here is the twist. Each shelf is quick to spot the other's jaw and slow to spot its own, and each, looked at closely, uses both. The evolutionary shelf commits the moralistic fallacy whenever it waves away the German border, which moved a whole population of mothers by 350 hours a year, because a culture that strong is inconvenient. The constructivist shelf commits the naturalistic fallacy whenever it argues from "the brain is a mosaic" to "therefore we ought to arrange things as if it were", which is Hume's forbidden step taken backwards.

So the naturalistic fallacy and its mirror are not two mistakes. They are the two jaws of one trap, and each shelf is standing in the jaw it cannot see.

Here is a shelf-one story taken to the archive, which I would have told you as fact a few years ago. Evolutionary psychology has long carried a hypothesis it calls the dual mating strategy: ancestral women secured provisioning from a reliable partner and, on the quiet, better genes from more formidable men. If so, the historical rate of children fathered by someone other than the husband should be substantial. It can be measured now. A team led by Larmuseau, working in the Low Countries, matched Y-chromosome analyses to genealogical records going back five hundred years, and published what they found in 2016 and again in 2019. Over those five centuries, well before modern contraception, the share of children fathered outside the marriage ran at about one or two in a hundred per generation. In tightly knit farming villages, where everyone knew everyone, it fell to fewer than one in two hundred. Among the poorest classes in the crowded, anonymous cities of the Industrial Revolution it rose to about six in a hundred.

Sit with that, because it is the cleanest thing in this chapter. A hypothesis that needed many cuckolds found few. And it found them where the neighbours had stopped watching. The bodies did not change between the village and the city. What changed was who was looking. That is a shelf-one hypothesis felled by data, and it fell in the shape this book keeps finding: a behaviour that exists, at a rate that rises and falls with the conditions around it.

The sentence that survives

If natural does not mean right, and unwelcome does not mean false, what grammar is left? The Preface gave it in one line: *given what we are, arrangement X under condition Y tends to produce Z.* Not "men are, therefore society must". A conditional, with the condition stated, and a verb that admits it describes a tendency in a population rather than a law for a person.

The careful authors on both shelves already write this way. Sarah Blaffer Hrdy, whose grandmothers come back later in this book, showed that the cost of raising a large-brained, slow-growing human infant could

never have been met by a mother and father alone, and that we evolved as cooperative breeders, with grandmothers and other kin doing much of the holding. She does not conclude that women ought to live in multigenerational communes. She concludes that the infant arrives expecting a network of caregivers, and that a society which isolates a mother in a flat with one child and one partner should expect the bill. That is a conditional. It does not prescribe.

So the conditional sentence does not just avoid the two jaws. It relocates the argument. Biology dictates the cost of a social arrangement, not its moral worth. Say a couple wants to split the first year of a child's life exactly in half. Nothing in the endocrine accent forbids it. What the accent does is set a price: the mother has paid the whole metabolic bill for nine months and, if she nurses, is still paying it, and the father's body, as a later chapter shows, is only now adjusting. A half-and-half year needs more scaffolding than the folklore assumes, and saying so is not the moralistic fallacy in reverse. It is a price list. Prices are not verdicts. You can pay them, and many couples do. You just should not be told they are zero.

Which gives you the question that replaces the old one. Not "is this nature or culture", which has no answer that survives the German border. Instead: which layer is this in, and what would it cost to change? That is the grammar for the rest of the book. Every recommendation from here on is a conditional or it is cut.

What held, what fell, what is still on the table

The Preface listed some of the fallen. This section is about telling in advance which way a finding is likely to go, because the fallen have a family resemblance, and so do the survivors.

Take the most famous fall and look at the mechanism. In 2010 a small study reported that two minutes of standing in an expansive pose raised testosterone, lowered cortisol and made people take more financial risk. It made a wonderful talk. In 2015 a replication in the same journal, with 200 people and the hypotheses written down in advance, found no effect on testosterone, none on cortisol, none on risk. One thing survived: people who held the pose said they *felt* more powerful. The original's lead author later disowned the finding in public, which is more than most findings get.

A small, unregistered study measuring a hormone, reported with a large effect that made a wonderful talk, is exactly the shape of thing that falls. The single-sex schools the Preface mentioned fell differently. Once you ask who gets sent to such schools, the advantage shrinks to a rounding error. A finding that depends on not asking who was in the room is the second family of the fallen.

The survivors share features too: large samples, many countries, effects that were big to begin with, and a physiology underneath. Upper-body strength, the Preface's different piece of equipment, holds in every occupational cohort tested. The interest gap holds, things against people, in every inventory it has been given. The fatherhood drop holds, measured in the Philippines and found again since. And the old argument between Janet Hyde's "most differences are small" and Marco Del Giudice's "the whole person is far apart" is not a failure of replication at all. Both sets of numbers reproduce. What differs is the question, one trait at a time or the whole profile at once, and the choice between those questions, I have come to think, is ideological on both shelves.

Then the middle, the largest category and the most useful. Give people a test of reading feelings from a pair of eyes and women do better. Pay men to be accurate and the gap goes. The best example of the middle is stereotype threat, because it cuts both ways and each shelf holds one edge. In 1999 three psychologists gave able women a hard mathematics test. Told it was known to produce sex differences, they did worse. Told it was not, they matched the men. It became the constructivist shelf's flagship. In 2012 Stoet and Geary went through the replications. Many reproduced the effect, and many of those had adjusted for prior mathematics scores in ways that manufacture it. Of the clean replications, three in ten worked. Both shelves misread that. Three clean studies in ten is a real phenomenon. It is also not what it was sold as, because a thing that happens in three rooms out of ten cannot carry a national gap in engineering. Negotiation has the same shape. Pool the bargaining studies and men come out with better economic outcomes. Have the women negotiate for someone else, or make the range transparent, and the gap disappears or reverses. Robust and movable are not opposites. They are the two halves of one conditional sentence.

The paradox, worked through

I promised one fight in full. This is the one I most want my shelf to win, so it is the one I should least trust myself on.

The pattern first, because the pattern is not in dispute. Study after study, from the early 2000s on, across tens of thousands of people in dozens of countries, has found the same odd thing. Women score higher on average in neuroticism and agreeableness, and the gaps are widest where sex roles are loosest. On risk, patience, altruism and trust, the gaps are near zero in poor countries and largest in rich ones. And then the one that made the newspapers. In 2018 Stoet and Geary, the same pair who had sorted the stereotype-threat replications, took the science scores of fifteen-year-olds in dozens of countries and found that in most of them girls matched or beat boys. Yet Finland, Norway and Sweden, the most equal societies on the index they used, produce fewer than a quarter women among their graduates in science and engineering, while Algeria, Tunisia and Turkey produce more than a third. The gender-equality paradox, it was called, and the newspapers loved it.

My shelf's reading is elegant. Where subsistence is not at stake, people follow their interests, and interests, on average, differ. Freedom reveals. I find it persuasive, and I notice that I do.

Now the challenge, which is longer than my shelf usually admits. In 2020 Richardson and colleagues, writing in the same journal, found that the graduate result rested on an undisclosed measure of women's "propensity" to graduate in those fields rather than on their plain share of graduates. Recompute with the share and the correlation shrinks, in some versions to nothing. A correction was issued. The same year Thomas Breda and his colleagues showed, on test scores from pupils in dozens of countries, that the stereotype that mathematics is not for girls is *stronger* in the richer, more equal countries, and that once you allow for it the link between a country's wealth and its gap in those subjects goes away. Then the study that matters most, because it asks what the cross-sections cannot. Fors Connolly, Goossen and Hjerm, also in 2020, took round after round of the same survey across Europe and asked: as a country becomes more equal over time, does its own gap widen? The cross-section reproduced. The within-country answer was no. The paradox, on this evidence, is a fact about which countries are which, not about what equality does

when it arrives. And in 2024 Herlitz and colleagues went back through the big studies and found that the paradox holds for personality and reverses for mathematics and partner preferences. Whatever is going on, it is not one thing.

So where does that leave it? The pattern is robust. Where societies free people most, some sex differences widen rather than close, and nobody yet knows why. My shelf's story explains the personality data and stumbles on the within-country test. The other shelf's explains the graduate data and has to work hard on the personality gaps, which are not small. The honest conditional, the only sentence that survives the above, is this: given that some differences widen where constraint loosens, and given that the stereotypes about them widen in the same places, anyone who wants to know which they are looking at tends to have to measure both, in their own population, over time, rather than read the answer off a map of Europe. That is less than I wanted. It is what there is.

That is where the fight stands, and when the paradox comes back later in the book it comes back as this: a real pattern, a contested cause, and a way of telling the two apart. That way is the tool, and it applies to my own findings as much as to anyone's. Here it is.

THE STRONGEST CASE AGAINST

The chapter has argued that the two shelves fail symmetrically and that a map of what held and what fell, read with a conditional sentence, is the way between them. The strongest case against is that the map is itself a shelf, drawn from my own side.

Start with the physiology the chapter leans on. Three decades of brain imaging, pooled, find that once you allow for the larger male head, sex explains next to nothing about the structure of a brain. The theory that testosterone in the womb organises the brain for life has been gone through study by study and its flagship natural experiment is confounded by years of surgery and medical stress that no control group shares. And testosterone in adults behaves as a result, not a cause: caring for an infant lowers it in men, competition and authority raise it in both sexes. On that reading the hormone levels of adults in a sex-stratified society are a *product* of the stratification, and a book that files them under physiology has put a cultural fact in the wrong drawer.

Then the behaviour, in Janet Hyde's voice, since she has spent a career counting. Most psychological sex differences are small, mathematics is a wash, verbal ability tilts the other way, and the chapter's "survivors" are the exceptions her literature already lists. Alice Eagly and Wendy Wood went further in 2023: most of the sex differences in the literature mirror what happens when you give people of either sex power, or take it away. If that holds, most of what the map calls a difference between the sexes is the psychology of the powerful and of the subordinate, learned before the child can sit up. The constructivist reply to the map is that what held is history, replicated.

And the sharpest point. Hyde warned that inflated claims of difference carry a social cost, because they become expectations. A chapter that offers a "price list" for equal arrangements is, on this reading, the naturalistic fallacy with better manners. It does not say the arrangement is wrong. It merely tells the couple, in advance and with a citation, that it is going to cost them, which is how the old shelf has always talked.

I have no rebuttal that closes this. I have the whole-person distance, the risk gap, the clean replications that did work and the wall that did not move the fathers, and they have the rest. That is a standoff, not a win.

IN PRACTICE

The scale here is you, alone, with a headline. None of this makes you right. It tends to make you slower, which in this subject is most of the job.

If the headline says "men are" or "women are", ask for the size of the gap first, and for the overlap in the same breath. A small gap is two bells sharing most of their area. A gap the size of grip strength is two different instruments. The failure mode is the gap on its own, which sounds precise and says nothing about the person across the table. The counter-evidence is the whole-person argument from the last chapter: many small gaps taken together can separate the sexes far better than any one of them, so "all small" is not the end of the story either.

If the finding is about a hormone, ask whether the hormone was measured or inferred. The fertile-days literature counted days on a calendar and fell when someone drew blood. Power posing fell when people were measured by researchers who had said in advance what they were looking for.

The failure mode is the plausible mechanism, which is what a story has instead of a result. The counter-evidence is the fatherhood drop: measured, followed, replicated, and still a hormone.

If the finding is convenient for your shelf, double whatever scrutiny you were going to give it. This is the only rule that works against you. The failure mode is the feeling of recognition, which is indistinguishable from being right. The counter-evidence is that convenient findings are sometimes true, the things-and-people gap being convenient for my shelf and still standing.

If the study compares countries, ask whether anyone has watched one country change. The paradox holds across a map of Europe and did not appear inside any country over successive rounds of the same survey. The failure mode is reading a photograph as a film. The counter-evidence is that within-country change is slow and the surveys may not have run long enough, which my shelf says and which might be true.

If the word "natural" appears, ask which layer, and then ask the price. Measurable in the body, statistical in behaviour, or variable with the culture, and in each case, what would it cost to change. A sentence that does not say is moving a claim from one layer to another while you are not looking. The failure mode is answering "natural" with "so what", which is the moralistic fallacy, or with "so there", which is the naturalistic one. The counter-evidence is that the question is sometimes honestly hard: whether the interest gap is a body's accent or a culture's echo is the fight nobody has won.

And if, after all that, the headline survives, say it in a conditional and carry it. Given this population, this measure, this condition, this tends to happen. It is a less thrilling sentence. It is the one I could defend at my mother's table, and I have had to.

Put the paperbacks back. Neither shelf needs you. The aisle does.

The studies, figures and open questions behind this chapter are on the website.

PART TWO

The Campfire Contract

CHAPTER 6

Flooding

The chairs are the kind nobody chooses for a living room. A woman and a man sit facing each other across a low table, and on one finger of each is a clip that reads the pulse through the skin. Under the shirt, on the chest, electrodes. On the palm, two more, for the sweat before there is any sweat to feel. They have been asked to talk about the thing they most disagree on, and they have chosen money, as a great many couples do.

At first the two lines on the strip of paper in the next room are what you would expect of two people sitting down. Then he says a sentence that begins with "you always", and the pen starts to draw a steeper line. Hers climbs. His follows. She answers. Nobody has raised a voice, and the two lines keep climbing. A little short of the line something in him closes. He is still looking at her, but the look has gone flat and the answers get shorter. When a line crosses 100 beats a minute the people in the next room make a note, because past this point the conversation is no longer being held by the two people having it. It is being held by their bodies.

At rest most people are well below it. Nothing visible happens at 100. But when a person reaches it sitting still, across a table from someone they love, it is an alarm, and the alarm is older than the argument, older than the money, older than the two of them.

The couple in the chairs are one of 79. The year is 1980, and the two psychologists in the next room, John Gottman and Robert Levenson, do not yet know that they will follow these couples for fourteen years, or that the strip of paper is going to become the most quoted, most misquoted and most oversold instrument in the study of marriage. This chapter is

about what the strip measured, what it did not, and why the honest version is the useful one.

The number on the strip

Gottman and Levenson recruited their 79 couples in 1980 and fitted them with the sensors I have described. The strip recorded the time between heartbeats, the time the pulse wave took to travel from heart to finger, the conductance of the skin, which rises when the sweat glands open, and how much the body fidgeted. They published the first wave in 1985 and the long follow-up in 1999, under the title "The Timing of Divorce".

Gottman's word for what happens past the line was flooding. It is a claim about the body, and it is measurable. Above the line the adrenal glands have released adrenaline and cortisol, blood pressure has risen, and the parts of the brain that handle listening and weighing are being quietly stood down in favour of the parts that handle running away. Where exactly the line sits was always a rule of thumb dressed as a threshold, and a fit person's line sits lower, because their resting pulse does. But the shape of the claim holds, and it is the first thing to take from this chapter. A quarrel has a physiology. Past a threshold, the body cannot listen, whatever the mouth is still doing.

Here is the first sex difference, and it is a fact about averages, not about the man across the table. In the 1985 sample, men flooded sooner than women, stayed flooded longer, and took longer to come back down. From this Gottman built his account of the thing he called stonewalling: the flat look, the short answers, the retreat behind the newspaper. The stonewaller, in his reading, is not refusing to engage. His body has learnt that silence brings the pulse down, and it reaches for silence the way a hand lets go of a hot pan. The technical name is escape conditioning, which is the right name, and it carries a kindness that the word stonewalling does not. Men on average flood sooner and take longer to come down, and the withdrawal is very often the body protecting the marriage from what it would say if it stayed. Most of the stonewallers he saw were men, Gottman has said, though that is his impression from the clinic, and nobody has shown me the count.

The strip also fed the part of Gottman's work that everyone has heard of. From the coded conversations he named the behaviours that predicted trouble and called them, with a flair for marketing the field has never quite forgiven, the Four Horsemen: criticism, contempt, defensiveness and stonewalling, with contempt the most corrosive. He gave the world a ratio of good exchanges to bad, which you have probably met on a fridge magnet. And then the claim that made him famous. In 1998, and again in 1999, he announced that his model could tell, from a conversation on tape, which newlyweds would divorce, and get it right better than nine times in ten.

Now the other column. Richard Heyman and Amy Smith Slep published a reply in 2001, in the same journal, with a title that is itself the argument: "The hazards of predicting divorce without crossvalidation". The problem was not the horsemen. The problem was arithmetic. Gottman had built his model on a sample and tested it on the same sample, and any model with enough knobs can describe the data it was fitted to. Whether it could describe couples it had never seen was never asked. The famous figure rests on those 79 couples and, within them, on 21 divorces. When Heyman and Slep built a model of the Gottman kind and tested it the fair way, on couples it had not been fitted to, the prophecy came apart. Tested like that, it often did little better than chance.

They went further in 2002, taking couples from the community rather than the laboratory and testing the specific sex-typed steps Gottman had described: the husband refusing to accept influence from his wife, the wife's harsh start to the conversation, the failure to calm things down. Those steps did not replicate. The things underneath them did. Contempt, criticism, flooding and withdrawal remained solid risk factors for an unhappy marriage, in that sample as in every sample since. What fell was the choreography, the idea that the steps had a fixed sex. Heyman and Slep used the word equifinal, many roads to the same town, and that is the honest summary of the programme. Gottman found risk factors. He sold prophecy. The idea that a few minutes of tape tells you who divorces is folklore. What survives the fair test is the risk, and what a couple does about it, which is the part of the work nobody put on a magnet.

Who wants the change

The second thing on the strip was not a number but a shape, and it took a different psychologist to name it. Andrew Christensen, in Los Angeles, described in 1988 the pattern he called demand-withdraw: one partner raises a problem, presses, criticises, asks for change, and the other goes quiet, changes the subject, leaves the room. The more one pushes, the more the other retreats, and the retreat is the thing that makes the next push harder. You can see the pulse inside it. The withdrawer is often the one past the line.

How big is it, and who stands where? The largest pooling of the studies, published in 2014, found what the folklore expects and then something it does not. The pattern goes with an unhappy relationship, reliably, in every kind of sample. It is more often the woman asking and the man leaving. And when it runs the other way, the husband pressing and the wife going quiet, which it does in a substantial minority of couples, it hurts at least as much. That rules out the comfortable story that it hurts because women ask. Demand and withdraw is a loop with two sides, and neither side is the fault.

So why is it more often her asking? Christensen compared couples in Brazil, Italy, Taiwan and the United States in 2006 and found the same asymmetry in every one of them. That looks like a body story. But put heterosexual, gay and lesbian couples through the same procedure and the demander turns out to be whoever wants more change in the relationship, regardless of sex. Strip the sex out of the pairing and the pattern survives, attached now to the person with the complaint. The partner who wants the house to be different pushes. The partner content with the house as it is withdraws, because the partner who wants nothing changed wins by saying nothing. In the heterosexual majority it is more often the woman who wants something changed, and chapter 7 says what. This is the one place in the chapter where I need the book's three layers. The pulse is physiology, and it is measured. The retreat is behaviour, and it is an average. Who demands and who withdraws is culture, and it is ours. Biology provides the mechanism, which is flooding. Culture decides who demands and who withdraws.

Let me say plainly what this book assumes. The heterosexual couple is its central case because that is where the endocrine difference and the cul-

tural script sit in the same bed. Same-sex couples are the control group. Where they show the same pattern, the pattern is not about the sexes. Where they do not, it might be.

The control group's own strip comes from the same laboratory. Gottman and Levenson ran the same protocol on a few dozen gay and lesbian couples, followed them for years, and published in 2003. The same-sex couples opened their disagreements with more affection and humour, showed less belligerence and less fear, and the unhappy ones among them showed oddly low arousal in conflict rather than the high arousal of unhappy husbands and wives. A few dozen is a small number, and so says Michael Rosenfeld at Stanford, who runs a nationally representative survey of how couples meet and stay together. His 2014 analysis found that once you allow for whether a couple is married, which for most of the period same-sex couples were not permitted to be, same-sex and different-sex couples break up at nearly the same rate. The control group floods less and splits at the same rate. The pulse is not the whole of it.

The sex that reacts more depends on the measure

Everything so far has measured the heart. Janice Kiecolt-Glaser, at Ohio State, measured the glands, and the sex difference turned around.

Her 1996 paper asked in its title whether men were really more physiologically affected by marital conflict than women, and the answer on the hormones was no. During and after a quarrel, women showed larger and longer-lasting rises in cortisol and adrenaline than men did, and the rises were largest when the husband withdrew. Put the two laboratories side by side and you get a sentence to carry through the book. On the heart, men react more. On the adrenal glands, women do. The sex that reacts more depends on the measure, and on who has just left the room. Which means that the silence a man reaches for to bring his own pulse down is, on average, the thing that sends his wife's glands to work. It protects him. It costs her.

Then she did something that nobody who reads the paper forgets. In 2005 her team raised small blisters on the forearms of a few dozen married couples by suction, and then had each couple argue. The blisters of the couples who argued with hostility healed at about 60 per cent of the rate

of the others, and their blood carried more of the signalling proteins of inflammation. A bad argument is a small wound that keeps other wounds open. That is measured, on the skin, whoever signed what.

The hormone of this chapter is cortisol, which chapter 3 introduced and chapter 4 measured on a roll of cotton under the tongue, so I only remind you of its Endocrine Accent, chapter 3's phrase for a hormonal tilt that colours a response without choosing it. What the couple laboratories add is that the accent is spoken differently by the two sexes and heard differently by their partners. The oxytocin spray that was supposed to make everyone kinder has been tried on couples before a conflict discussion, and the Preface has told you how the cuddle hormone fared under replication. The more interesting result from that work was that the same spray calmed the stress system in women and roused it in men, and in men the rousing went with warmer talk. One molecule, two sexes, two directions, and in neither a verdict.

How much does any of it matter over a life? The biggest pooling of the health studies finds that the quality of a marriage predicts health and mortality in the same small way that diet and exercise do. Small is not nothing. And the most domestic finding of all comes from a German cohort study: married men who described themselves as lonely had a flatter daily cortisol curve, the signature of a stress system that has stopped resetting at night. A marriage is not a shelter by being one. It is a shelter when the person in it does not feel alone.

The ledger

One pressure on a couple does not come from the partner at all. It comes from the payslip, and it is the cleanest demonstration in this book of a cultural line with a bodily readout.

A team with the pharmacy records of more than 200,000 married Danish couples, which in Denmark you can have, lined the couples up by the share of the household income the wife earned. Right at the point where her income passed his, prescriptions for erectile dysfunction medication rose, by a tiny amount at a very sharp edge, and the wives at the same edge collected more prescriptions for insomnia and anxiety. The design compares couples just either side of the line, who differ in nothing but which

side they are on. Two things tell you which layer it lives in. There was no jump among cohabiting couples. And there was none among couples in which the wife had already out-earned the husband before they married. The pill is physiology. The line is a contract, and it only bites the men who signed the old one.

The line shows in behaviour too. When a wife just out-earns her husband, the couple becomes more likely to tell the census something different from what it told the tax office, he rounding himself up, she rounding herself down. Nobody does that for a line that does not exist for them. Men entirely dependent on a wife's income are unfaithful more often than women in the same position, on the survey evidence, though most of those men, in any given year, are not. The curve is real and most men on it are faithful.

Does any of this end marriages? The longest-running American family panel asked what predicted divorce in two eras, with the line drawn at 1975. For couples married before it, a household that split the housework evenly was more likely to divorce than one in which the wife did most of it. For couples married after it, housework lost its power to predict anything. What replaced it was whether the husband had a full-time job: without one, the couple was about a third more likely to divorce in a given year. The wife's income made no difference either way. I was born in 1975, the year that line falls, and my parents' household was the kind in which the old contract still held. The script changed around the children of that year while the men it bound still worked under it.

Is the script universal? Across a few dozen countries, a husband's unemployment raised the risk of separation only where the breadwinner norm was strong. Inside one country, a wife out-earning her husband cost life satisfaction in West Germany and nothing in the East, where she had been expected to work since long before the wall came down. The pills are physiology. Everything else in this section is culture. Nothing in between is a sex difference. It is a difference in who was promised what.

The tide and the bond

Two more things end couples, and the evidence is a good deal less sure about them than the headlines are.

The first is desire. In the largest pooling of the studies, men on average want sex more than women, a difference that shrinks when you allow for men overstating and women understating, and does not vanish. In a Hamburg sample, women's stated desire fell with the length of the relationship and men's did not, though that study compared different women in different relationships rather than following anyone. Then the other column, which the authors of the big pooling printed themselves. Women on average get less out of sex: in heterosexual encounters men reach orgasm far more reliably than women do, and women carry more pain and more fear. The authors conceded that this almost certainly depresses what women say they want, and held that the measured difference exists whatever its origin. Both are true. They are not a contradiction. They are the subject. For the couple in the chairs the practical findings are two. The gain in well-being from more sex levels off quite soon. And the total desire in a couple matters more than how well matched it is.

The second is attachment, which is where the couple literature keeps its one well-measured sex difference in temperament. Men score a little higher on avoidance and a little lower on anxiety. The difference is there, it is small, and it depends on who answered, since in web samples it disappears. It also moves with age. In a study that tracked the same people for most of their adult lives, anxiety peaked in young adulthood and avoidance fell in a straight line thereafter. The avoidant man of thirty is, on average, a less avoidant man of sixty. Sometimes the same marriage gets to find out.

What, in the end, holds? In 2020 a large team put dozens of datasets and thousands of couples through machine learning, and the result is the most useful number in Part II. Everything about the two individuals combined, their personalities, their attachment styles, their sex, explained about a fifth of how good people said the relationship was now. The relationship itself explained far more. And for how a relationship changed over time, nothing in either set explained much at all. The single strongest predictor was whether a person believed their partner understood them and responded to them. Not sex. Not testosterone. Not the pulse. Whether, when the pulse climbs, there is someone across the table who notices.

Which brings me back to the chairs. Everything here was measured in a laboratory, and you do not have one. But you have a wrist. The pulse is the

one instrument in the whole couple literature that a person can read alone, in the middle of the argument, without anyone's permission. Count it. That is the tool.

THE STRONGEST CASE AGAINST

The case against this chapter is that I have told a body story about a contract. Here it is, made as strongly as I can, and its best voice is Christensen's own.

Put heterosexual, gay and lesbian couples in the same room and the demander is whoever wants change. Sex drops out of the equation. Add Heyman and Slep's community sample, in which the sex-typed steps, the husband refusing influence and the wife starting harshly, failed to appear while the sex-neutral risk factors held. On this reading flooding is a human response to being asked to change. The person asked is more often the man because the person asking is more often the woman, and she is more often the one asking because she is more often the one with the complaint, which chapter 7 explains. Nothing endocrine is needed beyond the fact that people have adrenal glands.

Then take the ledger. The Danish pharmacies show the clearest bodily readout in the chapter, and it vanished in cohabiting couples and in couples where she had always earned more. If men's bodies were responding to being out-earned as such, the cohabiting men would have been at the pharmacy too. They were not. They were responding to a promise, and a promise is a sentence, not a hormone.

Then the control group. Rosenfeld's representative data say that same-sex and different-sex couples, once you allow for marriage, dissolve at the same rate. If the pulse were destiny, the couples that flood less should last longer. They do not, measurably. And the machine-learning result gives everything about the individuals, sex included, a fifth of the present and almost nothing of the future. On this reading the Overlapping Bells are not a qualification to the chapter. They are the chapter, and the sex difference is a thin stripe on a map that is mostly water.

I think the case is largely right, and I have written the chapter so that it can be. What it does not do is make the pulse go away. Kiecolt-Glaser's blisters healed slower whoever had signed what. Flooding is the mechanism whichever culture is operating it, and a mechanism you can count

on your wrist is worth more to the couple in the chairs than a norm they cannot see. The contract decides who floods. The flood decides what happens next.

IN PRACTICE

Everything here is conditional, because everything above was measured on averages, and you are not one.

If you can learn to notice your own pulse in an argument, you have the one tool from the laboratory that works without it. Two fingers on the wrist, fifteen seconds, multiply by four. If the figure is around 100 and you are sitting down, the conversation has moved from your mouth to your glands, and nothing said until it comes down is likely to be heard. The failure mode is that counting is itself a withdrawal. A partner who sees you looking at your watch mid-sentence may reasonably take it as the newspaper by another name. So say what you are doing, before you need to.

If you are the one who goes quiet, the clinicians' rule of thumb for a pause is about half an hour, long enough for adrenaline and cortisol to clear and the pulse to come down, and it is a rule of thumb, not a measurement. The failure mode is the pause that never ends. A time-out that is not followed by a return at a stated time is stonewalling with a clock on it, and the partner left in the room, whose adrenal glands are now doing the reacting, is being flooded by the silence. The pause has to carry a time, and the time has to be kept.

If you are the one who pushes, consider that the loop is held in place by both ends and goes with unhappiness whichever of you is doing which. The question underneath it, in Christensen's reading, is who wants change, and it can be put on the table as a question rather than a complaint. "I want this to be different and you do not" is a sentence a man whose pulse is still short of the line can hear. The failure mode is that naming the asymmetry does not resolve it. The person who wants nothing changed still wins by waiting, and no amount of good phrasing moves a partner who has decided the house is fine.

If you keep score, Gottman's famous ratio is a serviceable shape for a week and a terrible one for an evening. The failure mode is the ledger itself.

A partner who hears "that is the third negative today" is being audited, and audits are contempt in a tidy shirt.

If you are a man whose wife earns more than you, and it bothers you, the Danish data say the bother is real, measurable and most likely to be found in men who married under the older contract. The practice is to say which contract you thought you were signing, since the jump at the pharmacy vanished in couples who had settled the question before the wedding. The failure mode is the opposite case. If it does not bother you, there is nothing here to fix, and importing the problem because a chapter said men have it is a way of acquiring a norm you were lucky enough to be born without.

If desire has fallen on one side, the finding that the total matters more than the match suggests the first question is not "why do you want less" but "what would make either of us want more". And if sex is reliably less rewarding for one partner, which on average it is for women, then the fall in her desire is information about the sex, not about her. The failure mode is scheduling as duty. Responsive desire needs something to respond to, and a diary entry is not it.

If one of you is the avoidant one, the averages say it is slightly more likely to be the man, and the long studies say it tends to ease with age. The practice is to read avoidance as a weather forecast rather than a verdict. The failure mode is patience as a substitute for change: a line falling over decades is a long time to wait, and the lonely men in the German cohort were married the whole time.

And if none of this fits, that is the Overlapping Bells at work. The husband who pushes and the wife who leaves are a substantial minority, not a footnote. Everything in this chapter is a risk factor, and the fair test is the thing to remember when a book, including this one, tells you what your marriage is going to do.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 7

The Second Shift That Wasn't

It is twenty past six in the evening and a woman is standing at a cooker in a flat somewhere in Germany with a wooden spoon in one hand and a pen in the other. The pen is the unusual part. On the table behind her is a booklet from the Federal Statistical Office that cuts her day into thin slices, and every slice has to be filled in: what she was doing, what else she was doing at the same time, who was with her. Stirring the sauce goes in the first column. Listening for the child in the bath goes in the second. The pen hovers over the slice from six to ten past, because in that slice she also found the swimming kit, answered a message from the nursery and remembered that the washing machine finished an hour ago. There is no column for remembering. In 2022 the Federal Statistical Office handed booklets like that to households across the country, and what came back is the closest thing Germany has to a ledger of how a household spends its minutes.

Try it yourself for one day. Just a sheet of paper with the hours down the side and a rule: write what you were doing, and what else. Most people who try it learn two things by bedtime. The first is that they do more than they thought. The second is that their partner does too, and that the two totals are closer than either had been prepared to admit at the previous argument. The third thing, which takes longer to see, is that the two columns do not hold the same kinds of minutes, and that a ledger which balances at the bottom can still be unfair on every line.

This chapter is about that ledger: who keeps it, what it says, where it lies, and what the body adds to it that the pen cannot catch. It is where

I have to tell you, in some embarrassment for my own sex and some for the other, what the diaries actually say about the famous second shift. It is not what either shelf expected.

The month that was not there

In 1989 the sociologist Arlie Hochschild published *The Second Shift*, a book built on interviews with and observation of a few dozen dual-earner couples, and gave the late twentieth century one of its most durable sentences. Women had gone out to work, she said, and come home to a second shift of housework and childcare, and the arithmetic of the double day added up, across a year, to an extra month of twenty-four-hour days. The revolution had stalled at the front door. My generation grew up inside the sentence, and it had the great advantage of all good explanations: nobody had to count anything to believe it.

Then people counted. The instrument is the time-use diary, and the whole chapter stands on it. Ask people how many hours a week they spend on housework and they estimate, and when they estimate they flatter themselves. A diary does not ask for an estimate. It asks what you were doing at ten past six, and then at twenty past, through the whole day, and what else, and who was in the room. Diaries like that have been collected across the rich world since the early sixties and pooled into one archive, more than a million diary days of them.

In 2019 Jonathan Gershuny and Oriel Sullivan, who have spent their working lives in that archive, summarised fifty years of it in a single review. The headline is the one Hochschild's sentence cannot survive. Add paid work to unpaid work and men and women do the same amount: about eight hours a day, for both sexes, and that total has barely moved in four decades. The extra month is not in the diaries. Where earlier studies found it, they were measuring an unusual slice of the population, and a few dozen couples chosen for a book are an unusual slice.

The second shift, as a gap in hours, is not there.

What the diaries do show is a different thing and, I think, a more interesting one. Inside the same eight hours, women do about three fifths of the unpaid work and men about three fifths of the paid. Unpaid work is converging, slowly: women do less of it than their mothers did, men a

little more than their fathers. And still, in every country in the archive, the routine tasks sit mostly with her. The total is equal. The kind is not.

Germany keeps its own ledger. The booklet from the opening, the Federal Statistical Office's survey of 2022, found women doing about an hour and a quarter a day more unpaid work than men. A decade earlier the gap had been wider, so it is closing, slowly, and the routine part closes slowest. Then the survey splits the sample by children. Mothers of children under six do about ten hours a week less paid work than women without children. Fathers of the same children keep their full-time jobs and trim their leisure a little. Which is how an equal ledger produces an unequal life, and how the double burden of my mother's generation in the West, the part-time job plus all of the housework, turns out to have been an arrangement rather than a law of nature.

The kinds of minutes

So why, if the totals match, do working mothers across the rich world say they are more exhausted, more short of time and more torn between work and home than working fathers do? There are two answers in the research, and both are about the texture of a minute rather than its number.

The first is density. In 2011 two sociologists took a different instrument to mothers and fathers in dual-earner families: a pager that went off at random moments through the day and asked what they were doing, what else, and how they felt about it. Mothers were doing two things at once for more of the week than fathers were, but the number is the smaller finding. The larger one is what the two things were. When a mother multitasked it was childcare on top of housework, and it came with distress. When a father multitasked it was more often leisure or paid work on top of something, and it did not. The same slice of the diary, "cooking", is a different slice if someone is in the bath upstairs. A diary that counts minutes cannot see the fragments, which is why the ledger balances and the woman holding it does not feel balanced.

The second is thought. In 2019 Allison Daminger published interviews with both halves of a few dozen couples, each partner on their own, and gave the invisible part of housework a structure. Every household task,

she found, has four stages. Someone has to anticipate the need before it is urgent: the shoes the child will outgrow by autumn. Someone has to identify the options. Someone has to decide. And someone has to monitor whether the thing was actually done. Deciding, the stage that carries authority and is visible to both partners and to the in-laws, was shared almost equally. Anticipating and monitoring, the stages that happen inside one head, continuously, with no slice in any diary, were done by the woman, even in couples who split the chores down the middle. He decides which summer camp. She noticed there was going to be a summer, and she checks that the form went in. Daminger called it cognitive labour. The internet calls it the mental load, and for once the internet is right.

Whether it hurts is a separate question with a first answer. One survey of mothers of young children found the thinking more skewed than the doing, and the skew tracked depression and burnout. One survey, and the arrow could run either way, but it points where the pager pointed.

What wears people down is not the number of minutes. It is the minute that can be interrupted, the minute that had to be anticipated, and the minute that has to be checked.

Now the German family panel, which the Preface quoted. You have seen its figures on laundry, cleaning and cooking. The figure you have not seen is the other column: men are primarily responsible for outdoor tasks and repairs in four couples out of five. Put the two columns side by side and the difference is not in volume, it is in rhythm. Laundry recurs, and cannot wait for a convenient Saturday. A dripping tap can, and usually does. The tasks that went to women are the ones that interrupt, and that is the oldest clause of the Campfire Contract from chapter 2: interruptibility, not strength, decided who did what around the fire. The modern couple has kept the clause and changed the fire.

The panel also shows when the clause arrives. Couples who begin symmetrical, both working, both cooking, almost invariably revert to an unequal split at the birth of the first child, and the Preface has already told you how many of the women in the households that result are thinking about leaving. The mechanism is in the German diary above: she drops ten hours of paid work, he trims his leisure a little, and the routine tasks fall into the hours she now has and do not climb back out when the paid hours return. Couples revert at the first birth, and they revert mostly by script. The body supplies one bottleneck, which is the milk. The script

supplies the rest.

And then the finding I keep turning over. Compare each man in the panel with himself over time, rather than with other men, and his life satisfaction goes up when he spends more hours on routine domestic work and fewer on paid work. He likes the laundry more than the breadwinning, it appears, once he does enough of it to find out. So why does the ledger not simply rebalance itself?

What the ledger costs

Because there is a second ledger behind the first, and it is kept in marriages and in beds.

Start with the marriages. The best study of this followed American couples across nearly half a century and asked whether the division of labour predicted divorce. For couples married before 1975, a wife in full-time work raised the risk, as the old economics of specialisation said it would. For couples married after 1975, the year I was born, it did not: a wife's job and her own income no longer destabilised anything. What predicted divorce in the modern cohort was the husband. A husband not in full-time employment carried a divorce hazard about a third higher. The requirement that she stay home has gone. The requirement that he earn has not. Which means that the man in the German panel who cuts his paid hours and feels better is, in the American panel, raising his odds of being left, and the two findings do not contradict each other. They are the two halves of the ledger.

And the bed, which the sociologists could not leave alone. In 2013 Sabino Kornrich, Julie Brines and Katrina Leupp found, in a large American survey from the early nineties, that couples with a traditional split of chores had sex more often than egalitarian ones. Three years later a second team took a survey from 2006 and found the sign had flipped: couples who shared the routine housework had sex a little more often than the traditional couples, and far more often than couples where the man did most of it. Two surveys, fourteen years apart, opposite answers, and the honest reading is that the script changed between them. Then the German family panel took the question apart by following the same couples over time. Hours did not predict sex. A fifty-fifty split did not predict sex.

What predicted it, for both partners, was whether she judged the split fair and whether he judged his own contribution fair. The ledger matters to desire. It matters through the sense of justice, not through the count.

Same-sex couples are the control group here, and they behave like couples who were never handed the script. They divide the work more equally than different-sex couples do, though not fifty-fifty: by time, by resources, by who minds which job less. The part worth stealing is not the split. It is that they talk about it, because nothing tells them in advance who does the laundry, and they call their arrangements fairer even when the hours are not equal. The research has a name for this, the fairness paradox: objective inequality predicts dissatisfaction poorly and perceived unfairness predicts it well. Good, because a couple can reach fairness without reaching equality. Bad, because a couple can also be talked into calling an unfair ledger fair, and the panel says that a good many women eventually stop believing it.

So much for the ledger the pen can keep. Now the part it cannot.

Two clocks in one bed

My own side of the argument expects its reward here, and I have to be careful, because the body does keep time, in both sexes, and almost everything that has been sold about it is wrong.

Start with the clock that is supposed to be famous. The textbook menstrual cycle has ovulation at the midpoint of a tidy month, and the textbook has now been checked against hundreds of thousands of cycles logged by the users of a tracking app. Only about one cycle in eight turned out to be the textbook cycle. The rest run longer or shorter, and nearly all the variation is in the first half, before ovulation. Which means a woman who ovulates early this month and late the next cannot be assigned a hormonal state by counting days on a calendar, and most of the older research that found cycle effects on the mind did exactly that. When researchers measured the hormones instead, a great deal went quiet.

Taste went quiet first, as chapter 5 told: the fertile-day preference for the square jaw did not survive hormones in the saliva. What does move is appetite, which peaks mid-cycle, with the partner and otherwise. The volume rises. The channel does not change.

Then cognition, which is the one the self-help shelf lives on. The study I trust most tested women on working memory, divided attention and cognitive bias at four points of a cycle, with the hormones measured, and then did the whole thing again over a second cycle in the same women. In the first cycle there was a hint: working memory fell as progesterone rose. In the second cycle it was gone. The authors called the first result a false positive and found no consistent link between any hormone and anything they measured. It is the only study in the field I know of that replicated itself before publishing, and it replicated nothing.

And the body. The largest pooling of exercise studies across the cycle found performance in the days of the period lower than in the rest of the cycle by a sliver so thin that the interval around it includes zero, and the authors advised against training rules for whole populations. For the average healthy woman, then: the cycle is measurable in the blood, visible on the thermometer, and all but invisible on the stopwatch and in the memory test. It is not the finding my side wanted, and I am giving it to you anyway.

Two things survive, and they are specific. The first is sleep. In the second half of the cycle progesterone raises core body temperature a little and blunts the night-time fall, and women say they sleep worse in the last week before the period. That is measurable, and it is the one piece of cycle-aware living with a thermostat attached. The second is the minority for whom the cycle is no rounding error. Premenstrual dysphoric disorder, the severe and diagnosable form of the premenstrual week, affects about one woman in twenty. What causes it was settled by an experiment that everyone who shares a bed should know. A team at the National Institute of Mental Health switched off the ovaries of women with and without the disorder, using a drug that silences them. The symptoms vanished. Then they added oestradiol or progesterone back at ordinary doses. The symptoms returned, in the women with the disorder only. Years later the same group showed the last piece: after the first month of add-back the symptoms subsided while the hormones stayed high, so it was the change in level that set them off, not the level itself. Women with the disorder have the same hormones as everyone else. They have a brain that responds to the rate of change. Physiology, measurable, nothing a schedule can be built around, and everything a partner can stop taking personally.

Now the other clock, which my own sex prefers not to mention. Men

run on a day rather than a month. Testosterone peaks in the early morning and falls through the day. What the daily tide does reliably is set a time of day, and the two sexes' preferred hours do not agree. Ask people when they most want sex, as one team did with several hundred of them, and the women, whatever their chronotype, peak in the evening. The men are more often ready in the morning. Only the women who are strong larks show a second peak after breakfast, so the morning tide and the evening preference meet comfortably in a minority of beds, and the rest negotiate. Whether the women's evening peak is biology or the shape of the working day, nobody yet knows.

Underneath both of those clocks is a shared one, and it may matter more than either. Till Roenneberg and colleagues in Munich built a questionnaire that does not ask whether you are a morning person but when you actually sleep on days you choose, corrected for the sleep debt you ran up on days you did not. Chronotype, it turned out, is a bell, with most of us near the middle, and it drifts late through adolescence before advancing again for the rest of life. Roenneberg named what happens when a late chronotype meets an early school or office: social jetlag, the equivalent of flying several time zones west every Friday and back every Monday. In a couple the question is not who is right about bedtime. It is how far apart the two clocks are. Similarity in chronotype predicts how satisfied couples are with each other, their sex life and their sleep, whether they are dating, newly married or old. And sharing the bed itself, on the one small test of it, helps rather than hurts, restless partner or not.

So here is the tool. The second shift as a gap in total hours did not survive the diaries. What survived is not the number of minutes but their kind: the routine against the episodic, the anticipated against the decided, the interrupted against the whole. The body adds two clocks that are real. A monthly one, for most women too small to schedule and for a few too large to ignore. And a daily one that both sexes run on, together or badly. The cycle is measurable. What it does to the ledger is statistical, and small. What a couple does about the ledger is theirs.

The only way to see the ledger is to keep it. Keep it by kind, for a week, with a column for the thinking. Then look at the clocks. Then decide, together, which lines you want to balance and which you are content to leave unequal and call fair, because the panel says you can do either, and that it is the not deciding that costs.

THE STRONGEST CASE AGAINST

The chapter has argued that the kinds of minutes carry a cost the totals hide, that the first birth is where the ledger tips, and that two bodily clocks are worth knowing about. Here is the best case that it is wrong on all three.

On the first, the cost of kinds rests on thin instruments: a few dozen interviews, one survey of mothers that cannot say which way the causation runs, one pager study in a single sample. None of it has been measured in the big diary archive. Set against it, the diaries: more than a million days, equal totals, and the routine housework falling sharply for women across four decades. On that reading the second shift fell, the convergence continues, and the mental load is a name for the residue. Gershuny and Sullivan would say it in fewer words.

On the second, the reversion at the first birth looks like body and is script, and that is the opponent's point, not mine. West German wives whose income passes their husband's do more housework the following year. East German wives, same tax code since reunification, do nothing of the kind. If the body's lilt set the ledger, the border could not. The body follows the ledger. It does not write it.

On the third, the two clocks are one clock and a rounding error. Measured properly, the monthly one moves muscle by nothing and memory by nothing that survives a second cycle. Partners cannot even detect fertility. The sex-timing data are questionnaires, the shared-bed data are a handful of couples in a laboratory, and nobody can say whether the women's evening peak is endocrine or the timetable. The honest instrument for a household is the pen, and the honest conclusion is the sociologists': what predicts satisfaction and sex is perceived fairness, which no hormone sets and no diary measures.

I cannot answer all of that. The second and third objections prove the chapter's layer rule rather than defeat it: the reversion is culture, measurably, and the clocks are physiology, measurably small, and the mistake this book exists to avoid is filing either under the other's layer. The first objection stands as stated. The mental load is a real thing with a thin literature, and I have told you how thin.

IN PRACTICE

For a couple, then. In conditionals, with the places where each one fails.

If you want to know what your household actually does, keep the diary, both of you, for one ordinary week, and add the column the official booklet lacks: what you were holding in your head. Write the kind of task, not only the time. The failure mode is that the ledger becomes a scoreboard, and the scoreboard is not what predicts satisfaction, fairness is, so read it once, together, and put it away. The counter-evidence is the fairness paradox: a couple can keep a perfectly balanced book and still feel cheated, and the reverse.

If both of you intend to keep your paid work after the first child, decide the laundry before the birth, not after, and decide who anticipates and who monitors as well as who does. The failure mode is that the decision made in month seven of the pregnancy is overruled in month three of the baby by lactation, which is the one bottleneck in this chapter that no agreement moves. The counter-evidence is the East German mother, whose body tips the same way and whose ledger does not, which tells you that the rest is negotiable.

If your division is unequal in hours and you both call it fair, it probably is, for now. The failure mode is in the panel: a good many women in over-burdened households eventually stop calling it fair and start thinking about leaving. If you are the one carrying the anticipating and the monitoring, say so in those words. Daminger's finding is that the deciding gets shared because it is visible and the rest is not shared because it is not. Make it visible: the list on the wall, the calendar both can see, the agreement that whoever notices the shoes also buys them.

If you are the man and the paid hours are the thing you could cut, the German panel says you would tend to be happier for it. The failure mode is the other ledger. In the American panel that husband carried a divorce hazard about a third higher. The condition, then, is that both of you reject the breadwinner script before the hours change, out loud, because the script is enforced by the partner and the in-laws as much as by anyone.

If she has a cycle, and she does, do not schedule the household around it unless she asks you to. For most women the measurable effect on mind and muscle is too small to plan for, and "cycle syncing" is a product. The failure mode is that a partner who has read this chapter starts explaining her cycle to her, which is the benevolent version of the old insult and lands

the same way. What does hold: the last week before the period tends to bring worse sleep and a warmer body, so a cooler bedroom in that week costs nothing.

If she has premenstrual dysphoric disorder, or you suspect it, treat it as physiology, because it is. Two cycles of daily charting gets the diagnosis, the right treatment acts within a day or two, and the cause is a brain that responds to hormonal change, not a character or a mood. The failure mode is on both sides: hers, in self-diagnosing from memory, which inflates the figures, and his, in reading a diagnosable disorder as a monthly excuse. The counter-evidence is the base rate: about one woman in twenty, so most couples arguing about the premenstrual week are arguing about something else.

If your clocks differ, measure the difference before you fight about it. A couple an hour apart in chronotype has a different problem from a couple four hours apart, and the gap matters more than the direction. The failure mode is the working day: social jetlag is imposed by the timetable, not by the partner, and a couple cannot fix what the timetable imposes, only share it.

And if none of this fits, which for some couples it does not, then the overlap is doing its job. Your household is one line in the ledger, and the ledger was never going to tell you what to want. What it can tell you is where the first birth tips it, and that is the next chapter's business: the father, and the tide that goes out when he picks up the child.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 8

The Father's Falling Tide

A man is lying on his back in a tube that is louder than anyone warned him, with a mirror angled above his eyes and, in the mirror, a photograph of a baby. He has been told to keep still. For a few seconds the baby is a stranger's. Then it is his own, six months old this week, and whatever moves in him when the picture changes is what the magnets are there to catch. Before the birth he lay in the same tube and had his head mapped for the first time. The two maps are about to be laid over each other, and in a few places the second has slightly less in it than the first.

He is one of a small group of first-time fathers in a study run by Darby Saxbe, and what she found in him is the middle of this chapter. What I want you to notice now is simpler. Something has gone out of him between the two scans, the way water goes out of a harbour in the afternoon. The harbour is still there. What has changed is how much of it you can see, and what the going-out has uncovered.

If you have had a child, or stood beside someone having one, you have lain in that tube without the tube. The Preface gave you the headline: a man's testosterone falls when he becomes a father who holds the baby. Chapter 3 gave you the figure from Cebu. This chapter owns the rest: what the fall is answering, what the mother's body gives up in the same months, what happens to the tenth father, and what the other clock in the house is doing while the couple decides whether to do this at all.

Three layers, once, and then I stop labelling. The hormones and the grey matter are measurable. What mothers and fathers do with a toddler on a staircase is statistical, two bells that mostly overlap. Who takes the

leave is culture.

What the hormone answers

Lee Gettler's young men in Cebu, in the Philippines, are the study everyone cites. He and his colleagues sampled their saliva in 2005 and again a few years later, by which time some of them had become fathers, and the fathers' testosterone had fallen by about a quarter to a third. What this chapter owns is the mechanism, and the mechanism is not a leak.

Three details of the Cebu design carry the weight. The men with the highest testosterone at the start were the ones most likely to have found a partner and become fathers by the follow-up. The fall came after. And the men doing three hours or more of childcare a day had the lowest readings at the end. Read those together and the hormone is not being drained by fatherhood as a status. It is answering what the man's hands are doing. The fall tends to follow the hours, so a father who is never in the room tends not to get it. Before the child, the hormone is doing mating effort. After it, parenting effort. The two trade against each other and the body keeps the books.

Fatherhood changes a man's body, and the change follows the caring. It does not cause it.

The replication I find most persuasive is not a laboratory. Among the BaYaka foragers of the Congo Basin, Gettler and his colleagues found that the men their neighbours rated as the more reliable providers and the more generous sharers had lower testosterone and higher oxytocin than the rest. Different continent, different economy, same direction.

Testosterone is not the only thing moving. Fathers answer the cues of an infant with raised prolactin and oxytocin, and the two track different things: a father's oxytocin goes with the fine-grained exchange of faces and sounds that the researchers call synchrony, his prolactin with the coordinated play in which he steers the child towards the world. And nobody has yet found a baseline difference in oxytocin between new mothers and new fathers. What differs is what the hormone is doing while it is there, which turns out to be the whole subject.

Two routes into one network

In 2014 Eyal Abraham, Ruth Feldman and their colleagues put first-time parents into a functional scanner and measured their oxytocin. The parents came in three kinds, by design. Mothers who were the main carer. Fathers whose partner was. And gay fathers raising an infant as a couple, with no mother in the house, who were therefore main carers by circumstance. The question was whether the parental brain belongs to a sex or to a job.

The answer came in two networks. The first is old and low, the amygdala and its neighbours, the circuit that makes a cue matter: vigilance, salience, reward. The second is cortical and newer, the circuit for working out what someone else wants and planning what to do about it. In the mothers the first led, and its activity rose with their oxytocin and with how closely their behaviour tracked the baby's. In the fathers whose partner did most of the care, the second led. And in the gay fathers both ran high, the amygdala as high as the mothers', the cortex as high as the other fathers', with a connection between the two that the other groups did not show. Across every father in the study, the strength of that connection rose in a straight line with the hours he spent in direct care.

The brain did not ask who had been pregnant. It asked who was up in the night.

Now the mother, and the study that changed the field's manners. Elseline Hoekzema and her colleagues scanned women before they conceived for the first time and again after they had given birth, against women who stayed childless, first-time fathers and childless men. The mothers' grey matter had shrunk, symmetrically, in the regions that do social cognition. The change was so consistent that a classifier could tell every mother from every non-mother. Behind it ran an endocrine event with no equal in adult life, oestradiol and progesterone climbing through the pregnancy to levels the body never otherwise sees, then both off a cliff at birth. The regions that had shrunk were the regions that lit when the women later looked at pictures of their own babies, and the more a woman's brain had changed, the stronger her attachment to the infant scored afterwards. Six years on, the same women could still be told from the controls, and the changes still tracked attachment.

"Grey matter lost" is the phrase the newspapers used and it needs poli-

cing. Nothing in the data says the women lost ability at anything. The researchers' own word is pruning, the kind the adolescent brain does when it specialises, and the attachment finding points the same way. The mechanism is contested in the respectable way: other teams, scanning mothers in the weeks after the birth rather than across the pregnancy, found grey matter going up, and the reconciliation most of the field now accepts is two tides, a gestational ebb driven by the placenta's steroids and then a postpartum flood driven by the baby. Contested as mechanism. Not contested as fact.

Back to the man in the tube. Saxbe and her colleagues followed their fathers from the partner's pregnancy to six months after the birth and found what nobody had looked for because nobody thought it was there: cortical volume down and the cortex thinner across the brain, most of all in the precuneus. The fathers who had lost more responded more strongly to pictures of their own baby. The fathers who had bonded most with the unborn child, who planned to take more time off, and who then spent more hours with the infant, lost the most. So far it is the mother's story told smaller. Then the other column. Allowing for how they slept and felt before the birth, the fathers with the larger reductions reported poorer sleep, more anxiety, more distress and more symptoms of depression at six months. A mixed bag, and for once that is the right phrase.

That is the falling tide, in the body rather than the harbour. What goes out with it is a layer of insulation. What is uncovered is the child, and the rocks.

The brain changes too, in both parents, and it changes by use.

Two consolations, both tentative. The going-out seems to be the price of the attunement and not a separate injury. And in a very large British biobank, the more children a person has parented, the higher their functional connectivity in later life, in both sexes, in exactly the networks that ageing usually takes first. If that holds, the tide comes back.

The tenth father

Now the rocks, by name.

James Paulson and Sharnail Bazemore pooled every study they could find in 2010 and put depression in the months around a birth at one father

in ten. In the third to sixth month after the birth the rate was higher still, though fewer studies were looking in that window. Interviews found less than questionnaires, and American studies more than the rest of the world, which tells you the figure is softer at the edges than the headline. And a father's depression and a mother's travel together. It is one weather system over one house.

The father's version wears a different coat, which is why it goes uncounted. Depressed fathers are far more likely to be working long hours away from home and far more likely to describe themselves as isolated. Irritability, aches, the office. The standard screening questionnaire was written for mothers and misses him.

Whether it matters to the child was settled in Britain. A large cohort there measured fathers' depression in the first weeks after a birth and the children's behaviour three years later, and found the odds of emotional and behavioural problems roughly doubled, and of conduct problems in boys more than doubled, after allowing for the mother's depression and for the father's later depression. The father's state in the first weeks was an independent variable in the child's life.

One father in ten is depressed around a birth, nobody is screening him, and it shows in the child.

The mother's version is older, more thoroughly studied and about as common. Two evolutionary readings of it circulate and I carry both. One says postpartum depression is a bargaining move, an honest and costly signal by a mother who perceives too little support or too little viability in the infant, aimed at the father and the kin. The other says it is mismatch: the exhaustion is ancestral and the relief that used to arrive with it, the aunts and the grandmother, does not. The two readings agree on the only sentence I need. Human maternal psychology did not evolve for a flat with two adults in it.

Which is where the cultures come in, and where the German state did something I did not expect to find. Across unrelated societies the month after birth is fenced, the Chinese mother sitting the month while her female relatives feed her, the Latin American mother in bed for weeks with her sisters running the house. Germany's *Wochenbett*, the childbed, is the same idea on the public purse. State-funded midwives visit the home for up to twelve weeks after a birth and look at the mother's body, the baby's weight and the pair's state of mind. I grew up in that system without

knowing it was one.

And the couple, in one breath, because chapter 6 owns it. Parents are a little less satisfied with their marriages than non-parents, the drop arrives as a step at the birth, and childless couples slide down a slope to the same place. The child does not cause the fall. It brings it forward.

What a father is for

If the parental brain follows the job, the reader who fought for Emancipation 1.0 is entitled to ask, with some edge, what is left of mothers and fathers as distinct contributions. Here is what the pooled studies leave, and it is less than the shelf of fatherhood books claims and more than nothing.

Attachment first. The best pooled study of it found infants slightly more securely attached to mothers than to fathers, by an amount you could not see in a room, and a mother's sensitivity and a father's predict the child's security about equally. The old idea of the single mother at the centre has been replaced, by the people who tested it, with a plainer one: an infant is built to hold several secure attachments at once, which in a camp insures the child against any one adult's death, and in a flat with two adults leaves it insured against neither.

Then the one difference that survives, in a small way. A test called the Risky Situation puts a toddler with its father and introduces a stranger who keeps coming closer, then a staircase, then a toy the child is not allowed. Fathers who push the child to the edge of what it can bear and then set the limit produce what its designer calls an activated child. Fathers who never push produce an under-activated one, who later shows more anxiety, and fathers who push without limits produce the over-activated one, who later shows more impulsivity. Pool the studies and fathers do more of this than mothers, by a margin that leaves the two bells almost entirely on top of each other. Millions of mothers push harder than millions of fathers. The difference is real and it is a lilt, and rough-and-tumble play is its vehicle: not the harbour, the sea.

Language is the oddest and I like it best. In one study of rural families where both parents worked, mothers and fathers were recorded talking to their toddlers and the children's language was tested a year later.

Mothers talked more. Fathers used more different words, more questions beginning with who, what, where and why, and more requests for clarification, which in practice often amounted to "Huh?". A father who does not know the half-word makes the child say it again. The variety of the father's vocabulary predicted the child's language in its own right, after the mother's input, income and education had been allowed for, and the pooled studies since agree. The name for it is the Bridge Hypothesis: the father as the less accommodating listener who stands between the nursery and the world.

Now the objection, which I think is half right. The fun playmate and the demanding listener may be artefacts of a man who is at home less, who reaches for high-arousal play in his short window and does not know the child's half-words because he was at work when they were coined. Nobody has checked whether the Bridge Hypothesis survives in fathers who become the main carer, or whether their language drifts to the maternal profile. My bet, from the brain data above, is that it drifts. But a bet is not a finding, and the honest entry is that this difference lives in the behavioural layer and may belong to the cultural one.

One old frame sorts fatherhood into engagement, accessibility and responsibility, which is knowing the dentist is on Thursday. The first two have shifted towards fathers over the decades. The third has barely moved, and chapter 7 owns that ledger. The brain scans measure engagement. Nobody has scanned responsibility.

When the father is gone the children fare less well, and the obvious objection, for decades, was that the families without fathers were also the families without money. The studies built to separate the two find a direct effect, smaller than the raw comparisons but consistent, on finishing school, on social and emotional adjustment and on adult mental health. And the control group the argument needs exists: the children of gay fathers, raised with no mother in the house, come out within the normal range and on some measures above it. What a child needs is the whole network. The second parent's amygdala learns the job.

The other clock

Everything above happens after the decision. Most of the people this book is for are still inside it, and the decision is made against two clocks that run at different speeds in the two bodies.

A woman's ovarian reserve, the stock of follicles she is born with, peaks before her own birth and falls from there, and age alone accounts for most of the difference between one woman's stock and another's. By the end of her thirties almost all of it is gone. The population curve is rigid. The individual is not: the blood markers of reserve predict a woman's chance of conceiving in the coming months poorly, if she has no history of infertility. Low reserve is not sterility. It is a shorter runway of unknown length.

A team of demographers turned the curve into a timetable. For a nine-in-ten chance of two children without IVF, start by 27. IVF moves that date by a few years and no more.

The man's clock is slower and runs in a different currency. Sperm quality declines measurably from his thirties on and his odds of conceiving fall hard only much later. But sperm are made continuously, each round of copying with a chance of error, and the father's age turns out to be the main driver of new mutations in a child. The children of fathers past 45 run a higher risk of autism, of ADHD and of bipolar disorder than the children of young fathers, by multiples. A multiple of a rare thing is still a rare thing, so I give none of the multiples here, where they would read as prophecy.

Then the layers that decide whether the clocks are read. The economists who measure what a first child does to earnings put the long-run penalty for a mother in Germany at about 60 per cent of what she earned before, a figure the Preface used and chapter 13 comes back to. In Denmark, by the same method, it is about a third the size. A woman facing Germany's penalty rationally waits, and the wait is paid in the currency above. Germany has fewer children than Germans say they want, and most of the shortfall is third and later children who never arrive. The strongest sentence in the fertility literature is a demographer's: fertility collapses where women reach equity in the university and the labour market and meet the old inequity at home.

The earnings penalty falls on the mother. The care penalty falls on the

father who is kept out, and on the child who gets one parent's amygdala instead of two.

That is the tool this chapter hands over, and it is a tide table rather than a verdict. The hormones in the father fall with the hours and come back with neglect. The grey matter in both parents goes out with the attunement and uncovers the rocks. The ovarian clock is rigid in the population and unreadable in the person, the paternal clock slow and expensive. And the whole arrangement is run, in Germany, against a penalty that cuts a mother's earnings by more than half and a childbed that the state still pays a midwife to visit. Measurable, statistical, ours. Now the case against.

THE STRONGEST CASE AGAINST

The strongest case against this chapter is that it has smuggled a role difference across the border as a sex difference, and the chapter's own best study is the witness.

Abraham and Feldman's parents showed that the maternal network is not maternal. It is the primary carer's. The gay fathers raising infants without a mother lit the amygdala as strongly as the mothers did, and the connection in every father rose with his hours of care. Saxbe's fathers lost cortex in the same direction as Hoekzema's mothers, scaled to their hours. Oxytocin at baseline is the same in both parents. Attachment to fathers is less secure than to mothers by an amount that rounds to zero. Even the activation difference leaves the two bells almost wholly on top of each other, and Joseph Pleck's reading, that the father plays rough because he has a short window and does not know the child's half-words because he was at work, has never been refuted, only answered with a question nobody has run. On this view a chapter titled for the father's hormone has spent its length showing that the hormone is a thermometer for who does the nights.

The testosterone evidence is thinner than its fame. The Cebu fall is correlated with hours of care in an observational cohort, and the size of the sample and of the drop are given differently by different hands. Nobody has assigned fathers at random to nappies. The mother's brain is contested in direction, and the two-tide story that reconciles the studies rests on animal data and a review rather than on women measured through both tides.

And the deepest version. Across the ancestral populations that chapter 2 carried, a father's presence improved a child's survival in a minority of them, and a grandmother's in most. The record says fathers were useful, substitutable and optional, which is a strange foundation for a chapter about what a father is for.

I accept most of this. The plasticity is the finding, not an objection to it. What I keep is narrower. The second parent's brain learns the job, but it learns it with the hours, and in the countries this book is about the hours are not distributed by a coin. If the role is what matters, then the role is what we are arguing about, and the chapter is about who gets it.

IN PRACTICE

What follows is conditional, at the scale of a family, because that is all the evidence licenses.

If you are a father who wants the change the headlines promise, then the Cebu and Congo data say it tends to follow the hours of direct care and not the fact of the child. Three hours a day is where the effect was largest, which is a place on a curve and not a prescription. The failure mode is expecting the softening without doing the nights, or reading its absence as a verdict on your character rather than on your timetable.

If you are the partner of a new father and you want the second amygdala in the house to learn the job, then the straight line in Abraham and Feldman's scanner says the connection grows with the hours he spends alone with the baby, not beside you. Which means the sink. The bath, the bottle, the night that is his because the baby is his, the pram out of the door without a briefing. The failure mode is the gatekeeping both sexes do without noticing: handing over, hovering, taking back. Nobody has yet scanned a father who was handed the baby and left, so this is the direction of the evidence, not its proof.

If the baby sleeps in your room, then the night is shared by geography and his half of it is harder to hand off. If it sleeps elsewhere, decide whose feet hit the floor before the first night, not during it. The failure mode is the arrangement nobody chose, which tends to be the one where the mother wakes because she always has.

If you are deciding who takes leave, then the data on both parents say the one who takes it tends to be the one whose brain changes, and that the

change is not free. Saxbe's fathers who took more leave lost more cortex and slept less. The failure mode is treating the leave as a holiday and the one who takes it as the one who has it easy.

If you are a father in the third to sixth month and you find yourself at the office later than you need to be, or oddly short with everyone, then that is the window in which depression in fathers runs highest, and the symptoms in men are hours and irritability more often than tears. The failure mode is that nobody is screening you, so the only screen is the person across the table, and the counter-evidence is that the figure for that window rests on few studies.

If your partner is depressed after the birth, then you are at raised risk too, and the British cohort says your state in the first weeks tends to show in the child years later independent of hers. The failure mode is the division of concern in which one parent is the patient and the other is the staff.

If the mother in your house is being told she has lost brain matter, then she has, in the regions that now light for her baby, and so, at a smaller scale, has the father. The failure mode is the "baby brain" joke, a sex-difference claim dressed as sympathy, which the fathers' data now also disprove.

If you are two adults and no one else, then you are running an arrangement no camp ever tried, and the relief that was supposed to arrive does not. Build the substitute on purpose: the mother's mother if she is near, a neighbour's reciprocity if not, the state's midwife where there is one. The failure mode is pride, and its symptom is the sentence "we are managing".

If you are a couple deciding when, then the timetable is the only number in this chapter that speaks to you directly: for a nine-in-ten chance of two children, start by 27 without IVF, a few years later with it. The failure mode is reading the population curve as your own body, which the blood-marker data forbid in both directions. If the father in the plan is past 45, then the paternal-age risks are real and multiply small numbers. The failure mode is panic, because a multiple is not a probability, and the second is the dinner-party silence in which only the woman's age is a subject.

If you are in Germany, then you are inside the earnings penalty and the twelve-week midwife at the same time, which is the whole paradox of this book in one household. The penalty tends to delay the child, the midwife tends to buffer the birth. The failure mode is believing that either one is nature.

And if you take nothing else from the chapter, take the direction of the arrows. The hormone follows the hands. The brain follows the hours. The tide goes out on its own schedule, and what it uncovers is yours to walk on. A tide table is not a verdict.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 9

Who Keeps Children Alive

A woman of about seventy is standing at a sink in a kitchen that is not hers. It is late on a weekday afternoon, in a third-floor flat in a German city, and the child she collected from nursery is on the floor behind her, working a wooden train through the legs of a chair. She knows where the plasters are kept in this flat and where the spare key is. Her daughter's train gets in after five. Her son-in-law is in another city until Thursday. There is no rota on the fridge with her name on it, and if you asked her whether she was providing childcare she would look at you the way she would look at someone asking whether she was providing air.

Stand in the doorway a moment and watch whose hands do what.

The child falls, the train hits the skirting board, and the hands that pick him up are twenty years past their last baby and not noticeably slower for it. When her daughter comes in there is a handover of perhaps four sentences, two of them about the child's digestion, and what happened in this flat this afternoon appears in no statistic the city collects, because the city does not count it.

The Preface called a body which switches off its own reproduction and then lives on for decades a body that is for something other than the next child, and named that the *Grandmother Clause*. Chapters 1 and 2 found the clause in the whales and at the campfire. This chapter brings it indoors.

The woman at the sink is the clause, enforced at the scale of a flat. Nobody drafted her into it, and nobody can be sure that her doing it is good for everyone in the room, because the record on grandmothers is

not a sentimental one. It is a record of who lived and who did not, kept in parish books, pension rolls and the growth charts of children, and it contains both the most reliable helper the human child has ever had and a relative that one set of parish registers names, in the title of the paper, as the devil in the house.

Forty-five populations

In 2008 two anthropological demographers working in London, Rebecca Sear and Ruth Mace, published a paper with a title that reads like a child's question: who keeps children alive? Their method was plain and enormous. Find every population, historical or living, in which births and deaths were recorded well enough for a child's chance of surviving to be matched against which relatives were alive and nearby, and ask of each relative whether their presence made a measurable difference. They found 45 such populations, from parish registers to anthropologists' field notes, none of them with contraception and most of them without doctors.

The mother mattered in every population where her presence was tested. Her death was the single most reliable predictor of her child's. Older brothers and sisters helped in most. The mother's mother helped in about two in three. And the father, the figure around whom my mother's generation's whole argument about breadwinning was built, improved his child's survival in about one in three. In the rest, his presence or absence made no measurable difference to whether the child lived.

Across the human record, then, the grandmother is the person other than the mother who most reliably keeps a child alive. The father matters less than the folklore says, and more than nothing.

Let me say which layer that lives in, because it is easy to hear wrong, and it is the only time this chapter needs to. The finding is statistical, and it is statistical across populations, not across fathers. It says that in most of the pre-modern world, if a father died, the child's odds did not move, because other hands closed over the gap, and that if the mother's mother died, they did. It is a finding about which hands the human household was built around, and the number of those hands was not two.

What this chapter adds to the campfire of chapter 2 is what the record says when the camp becomes a farm, a parish, a pension roll and a flat.

The answer, it turns out, depends on whose mother the grandmother is. Sear and Mace's two in three was for the mother's mother. For the father's mother the result is not a fraction, because it does not point one way. Her effect was variable, and in some populations it was negative.

There is a reason to expect that. Human women live decades past their last birth, which almost no other mammal on land has thought worth doing, and the explanations compete. The grandmother hypothesis, which chapter 2 had at the digging stick, says the old woman fed her daughter's weaned children so the daughter could bear again sooner. A rival theory starts from the fact that in our species it was mostly the woman who moved to her husband's people at marriage. A young wife arriving in the house shares no blood with her mother-in-law, while the mother-in-law is kin to the young wife's children. If both breed at once they compete for the same food, and the older woman loses more by competing than by yielding, so she yields, and over enough generations the yielding became menopause. I don't know whether that is right, and the people who do this for a living have not agreed. Notice what it predicts, though: that the grandmother who yields is the husband's mother, and that she and her daughter-in-law are rivals by construction. The reader who is fond of grandmothers should brace.

The devil in the house

Krummhörn is a stretch of farmland in East Frisia, in the far northwest of Germany, and its parishes kept registers of births, marriages and deaths through the eighteenth and nineteenth centuries complete enough for families to be reconstructed across generations. Eckart Voland and a colleague went through tens of thousands of those entries to ask Sear and Mace's question one grandmother at a time. A maternal grandmother alive and nearby lowered infant mortality. A paternal grandmother living in the household raised the chance that a newborn died in its first month by two and a half times. The title the same two authors gave to their work on these registers says that the husband's mother is the devil in the house.

That is one parish. Here is a village in Mali. The Dogon are patrilocal, which means a bride moves to her husband's compound, and Beverly Strassmann, an anthropologist at the University of Michigan, followed

their children for years and found that a small child living with a paternal grandmother faced twice the risk of death. Dogon girls, she found, grew faster when their maternal grandmother was absent, and her reading is that the girls were spared the weeding they would otherwise have done for the old woman. Her larger argument is the one the grandmother hypothesis has to answer: in a household where calories are short, an old woman is not a helper but another mouth.

Here is a third country. A team working through the historical parish records of Finland tested every kind of relative they could name against child mortality. A living maternal grandmother cut a child's risk of dying of smallpox, and of diarrhoea, sharply. Paternal grandmothers and paternal cousins raised it. And the lethal effect of the father's side was strongest in the wealthiest families, which is not what a story about scarcity predicts. It is what a story about inheritance predicts. Where there was most to pass on, the husband's mother's presence was most dangerous to the children of the woman who had married in.

Three populations, three continents, the same sign. Ask living adults which grandparent gave them most care and the order does not vary: mother's mother, then mother's father, then father's mother, then father's father. The old explanation was paternity, since a mother's mother knows the child is hers and a father's father is unsure twice over. A stranger one puts the favouritism in the X chromosome, which a paternal grandmother passes through her son to his daughters and to none of his sons, and across several historical populations her presence did raise granddaughters' survival and lower grandsons'. The favouritism may be in the chromosome before it is in the kitchen. Nobody has watched the mechanism work.

This is the kind of finding that gets carried out of the room as a rule, so let me be exact. The paternal grandmother at the sink in the opening scene, if she is one, is not more dangerous to anybody in a city with vaccines and a supermarket, because the mechanism she was dangerous through was food and inheritance, not malice. The devil in the house was a hungry house.

The pension and the granddaughter

If the grandmother's effect runs through food, then give her money and watch where it goes. That is what South Africa did, by accident, in the early 1990s, and what Esther Duflo, the economist at MIT, did on purpose.

The country extended its old-age pension to Black households, at a level that was generous by the standards of the countryside, and paid it to the old person rather than to the household. Duflo asked what happened to the bodies of the children who lived with the pensioner. She measured weight for height and height for age, the two numbers a clinic uses to tell whether a child is being fed. When the pension went to a grandmother, her granddaughters grew, by more than a full standard deviation on both measures. Her grandsons gained nothing measurable. When the pension went to a grandfather, nobody gained anything, granddaughter or grandson. The same money, the same house, the same children, and whether the child grew depends on the sex of the hand that holds the cash and the sex of the child. Economists had modelled the household as a single purse. Duflo's finding is that it is not a purse. It is a set of relationships with money passing through them, and the relationships have sexes.

Money in a grandmother's hand reaches the child. Money in a grandfather's hand, often, does not.

That is the first of the two ways a state has of keeping a child alive while the mother is at work: it can pay the grandmother. The second is to run the nursery.

Now us. The grandmother who kept children alive in those 45 populations did it in conditions of scarcity, and most of my readers do not live in scarcity. So what does the clause do in a country with vaccines? Nobody has measured it as well as Duflo measured the pension, and the evidence has two halves that do not quite agree.

The first half says the household still runs on more hands than two, and suffers when it has only two. A survey of parents across dozens of countries, built to find out why parents burn out, found that the strongest predictor at the level of countries was not wealth, not inequality and not how much the mother worked, but how individualist the culture was. Where the extended family is still expected to turn up, parents burn out less, even when everything else is harder. The largest pooling of attach-

ment studies found that a father's sensitivity predicts his infant's security about as well as a mother's does. A child does not check the sex of the hands. What the infant is built for, on this reading, is a network of attachments rather than one, and where the network collapses, in families under severe strain, the most troubled form of attachment rises sharply. Children were never built for two carers, and are worst off with one.

The second half says the hands have to be chosen. Rebecca Sear's own later review found that living near relatives raised fertility in historical populations, and that in modern rich societies the effect of kin on a family is mixed or negative. Young adults forced back under their parents' roof by rents and insecure work are measurably less well in the mind for it. The extended household, in other words, is a buffer when it is voluntary and a cage when it is not. Among the Agta of the Philippines, where children and their carers wore proximity sensors for a season, kin gave care in proportion to the child's need and non-kin in proportion to what the other household gave back. Reciprocity predicted care as strongly as blood did. Nobody in that camp was trapped with anybody, and the neighbour counted as much as the aunt.

Then the nursery. The best pooling of studies on small children's stress hormones found that cortisol, which should fall across the day, rose from morning to afternoon at centre-based nursery and did not at home. That is measurable, it has replicated, and it is modest. What it means for the child in later life is not known, and the comparison my mother's generation would have wanted, a nursery against an isolated flat with one exhausted adult rather than against an idealised home, has not been run. In the forager camps, anyway, adults leave the frame early: children are taught by adults only while very small, and by other children after that. That is not an argument for any particular nursery. It is a reason not to assume that the two-adult flat is the baseline the nursery is departing from.

Put the two halves together. A mother's job does not hurt her child when someone else is in the room. It is the empty room that hurts.

The child on the floor

Now the children themselves, and the question my mother's generation thought it had closed: whether a boy and a girl arrive different or are

made different.

Start with the girls chapter 3 introduced, those born with congenital adrenal hyperplasia and exposed to adrenal androgens before birth, who as children choose trucks and blocks more than their sisters do. What belongs here is what the cameras saw in the playroom. The parents praised these daughters for touching dolls more than they praised their unaffected daughters, which is the opposite of the socialisation story, and the girls went on choosing the trucks, including when they played alone with nobody to please. The more severe the condition, the stronger the preference. By the first birthday the human preference, boys for boy-typical toys and girls for girl-typical ones, is among the largest behavioural sex differences anybody has measured, in the same league as height, and like height it overlaps. The girl with the truck and the boy with the pram are in every nursery, and nothing in a bell curve tells you which child in front of you they are.

And here is what the measurable layer cannot yet explain. At birth there is nothing. A famous study had newborn boys looking longer at a mobile and newborn girls at a face, and it went round the world, and a pooling of every such experiment since found no difference at all: both sexes prefer faces. Nothing on the first day, a large difference by the first birthday, and the field does not know what happens in between. I am leaving that open rather than filling it with a story.

The cultural layer has a surprise in it. The usual account of the toy aisle is that parents make it. The best test pooled every observational study of parents with sons and daughters, thousands of families, on two things parents demonstrably do, control their children and support their independence, and found that parents treat the sexes almost exactly alike. That does not mean culture is absent. It means that whatever is sorting the nursery into trucks and dolls, it is not mainly the parents' hands, and the reading I find plausible and cannot prove is that children sort themselves, choosing the games, playmates and models that fit a bias they arrived with. Marketing then sells to the sort.

Between the two sits temperament. The capacity to sit still, wait and do the dull thing first favours girls, by a wide margin in the older cohorts. The appetite for speed, noise and high-intensity pleasure favours boys. Sadness and fear show no difference at all. The gap widens through childhood and is at its widest in the years when a child is asked to sit at a desk all day, which

is why girls have been ahead in teachers' marks in every subject, by a small margin, in every sample since marks were kept. A teacher's mark rewards exactly the thing girls arrive with more of. The boy crisis is not a crisis in the grades. It is a crisis in what the grades are now worth.

The remedy that has been tested is the one the Preface listed among the fallen. Single-sex schools look wonderful until you ask who is in them, which is the selective, the private and the well-off. Allow for that and the benefit is the size of a good night's sleep. What the evidence does support is quieter and cheaper. A boy assigned to a male teacher scored higher, was rated higher and was more engaged, enough to close up to a third of the gap in English. A mentor who stayed more than a year lowered a boy's behaviour problems and raised his sense of his own worth, and a mentor who left within months did measurable harm. Neither works by teaching differently. They work by showing a boy that a man reads, and by not leaving.

Then adolescence, where the two sexes lose different things. The girl who argued as a child learns in her early teens that keeping her relationships means keeping her objections to herself, and the silence tracks the rise in girls' depression and eating disorders in the same years. The boys, interviewed at length by a developmental psychologist in New York, told her at thirteen that they loved their best friends, told them their secrets and would go to pieces without them. By seventeen the same boys had stopped. They had learned that the intimacy could be read as feminine or as gay, they put on what she calls a mask, and the loss tracks the rise in depression, isolation and suicide among young men. The scholar who first wrote about the girls' lost voice has since revised her book to say that the voice was never female at all, but human, coded as feminine by a culture and then taken from boys.

Here, then, is the tool this chapter hands over. At home the question is not whether boys and girls differ, which they do, not at all on the first day and a lot by the first birthday. Nor is it whether the parents made the difference, which on the best evidence they mostly did not. The question is which hands are in the room, whose mother they belong to, what the room costs, and what the room does with a child at the far end of one of the bells. The woman at the sink is the oldest answer to the first of those questions. The parish books say she is also, under the wrong conditions, the oldest problem. The clause does not say she must be there. It

says the household was never built without someone like her. The clause is a design, and the modern flat is the first household that has quietly removed the person it was written for.

THE STRONGEST CASE AGAINST

Take the chapter at its weakest, in its best voice. Beverly Strassmann's reading of her own Dogon data is that the grandmother hypothesis has been sold on the populations where it happened to work. Where food is short and the bride moves to the husband's compound, the old woman is a competitor for calories, her presence doubles a child's risk of death, and girls grow faster when she is absent. The Finnish records put the paternal side's lethal effect highest among the rich, which makes it an inheritance story, and inheritance is culture. The X-chromosome result is a correlation with a mechanism nobody has seen operate. Sear's own review says that in rich countries kin have mixed or negative effects, and the young adults pushed back into their parents' homes show that putting the generations under one roof in a modern economy costs people their peace of mind. In short: the grandmother who kept children alive did so in a world of smallpox and short rations, and the clause this chapter wants to carry into a flat with a fridge is a clause from a lease that expired. What the modern household needs is hands. The child does not care what sex the hands are, and the Agta show that need and reciprocity, not blood, predict who gives care. The Grandmother Clause, at home, reduces to "get help", and the word grandmother in it is sentiment.

And on the children: once brain size is allowed for, sex explains almost none of the variation in brain structure, the newborn study fell, and the finding that parents treat sons and daughters alike can be read the other way, as proof that the differences which remain are made by the agents nobody measured, which are peers, screens and shops. On this reading there are two small bells, one measured in trucks, and a culture that turns small gaps into separate schools and lonely boys.

I think that case is right about the flat with a fridge and wrong about the hands. The chromosome arithmetic is a correlation, agreed. Duflo's pension is not. The same money, in a grandmother's hand, grew granddaughters by more than a standard deviation and grandsons not at all,

and in a grandfather's hand grew nobody, and once the pension arrived I would not call that house a hungry one. Whatever filters help through the sex of the giver and the sex of the child was still working in the 1990s, with cash. That is the finding the case against has to explain, and it has not.

IN PRACTICE

This is the family's scale, so the you here is a household: two adults, or one, and the children, and whoever else is within reach of the door.

If the mother's mother is within reach, then the oldest arrangement in the record is available to you, and it tends to put more hands in the room at the end of the day. The failure mode is the one Sear found in rich countries: kin whose help comes with a view on how the child should be raised, so that each hour of relief costs an hour of argument. The counter is that proximity is not co-residence, and that the Agta kept their helpers by reciprocity, not obligation. Ask her what she wants back.

If the help on offer is the father's mother, then the parish books are not a verdict on her, because what they measured was food and inheritance and your fridge is full. The friction they predict is still there, in a different currency, and the Finnish records put it highest where there was most to inherit. Name the currency before she moves in. If she is to move in, then the co-residence data say it tends to go badly when money forces it, and the cohousing data say it tends to go well when each party keeps a door that closes.

If you are two adults and nobody else, then you are running an arrangement that appears nowhere in the 45 populations, and the individualism that makes it possible is the strongest predictor of exhaustion in it. The failure mode is to treat the exhaustion as a defect in the marriage or in one of you. The counter is the Agta: unrelated households that traded childcare did as well as kin. Find a second household and keep accounts, lightly.

If the child is small and the choice is a nursery or one adult alone in a flat all day, then the cortisol data say the nursery tends to raise an afternoon stress hormone a little, and say nothing at all about the flat, because nobody has measured the flat. The failure mode is to read a small, real effect as a verdict on working mothers, which is exactly what the language

of my mother's country stood ready to do. The counter is the forager record: adults leave the frame early, and children have been raised by other children for most of the time there have been children.

If you have a son who cannot sit still, or a daughter who chooses the truck, then the averages are real and far apart and your child is one of the many who are not at them. The failure mode is correction. The parents on camera corrected, and the girls went on choosing the trucks when they played alone. Letting him wrestle is not a decision to raise him differently. It is a decision to let him be where he already is.

If your boy is behind when the marks start, then the marks measure the capacity to sit still before they measure anything else, and girls have led on it in every sample since marks were kept. The failure mode is the separate school, which delivers a night's sleep once you ask who else attends. The counter is the matched classroom: a male teacher closed up to a third of the gap in English, and a mentor who stayed did more than one who left. Find him a man who reads and does not leave.

If your son is in his mid-teens and his friends have thinned to nothing, then the interviews say the thinning tends to happen at that age, is enforced from outside, and tracks the rise in young men's depression. The failure mode is to read his silence as the male condition. It was not his condition at thirteen. Ask about the friend by name. If your daughter has stopped arguing with you, then the failure mode is relief. The voice she is losing is a human one, and the boy down the hall is losing the same thing by another route.

If the father is the one at the sink, then the child does not know the difference, and chapter 8 has what his hormones are doing about it. And if you are the woman at the sink, then nothing in this chapter tells you what to want. It tells you that the flat was built on the assumption of your hands, that no one in it has counted them, and that the clause was always written in your favour and never signed by you.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 10

The Factor That Fell

You are sitting at a table with two people you have never met, in a university room with the blinds down, and a researcher has just put a sheet of paper between you. It shows a visual puzzle. Soon it gives way to a brainstorming task, then by a moral judgement with no right answer, then by a negotiation over scarce resources in which the three of you must divide too little among too many. Nobody has told you that the point of the afternoon is not the puzzles. The point is whether the three of you, taken together, have an intelligence of your own.

That is a strange thing to ask about a table. We are used to asking it about a person. Give one person a dozen different mental tests and the scores move together: someone good at one tends to be good at the others, and the thing they share, which statisticians long ago called *g*, is one of the most replicated findings in psychology. The question Anita Williams Woolley and her colleagues asked in 2010 was whether a group has a g of its own. Give table after table of strangers the same battery and see whether the tables that do well on the puzzle also do well on the negotiation. If they do, a group has a single underlying ability, call it c, and you can measure it, predict with it and perhaps build for it.

They found it. Or they said they had, which in the years since has turned out to be a different sentence.

What happened to it on that second look is the shape of this whole chapter. A team is where the sexes meet in numbers larger than a couple and smaller than a parliament. It is where my mother's generation fought for the seat, and what the seat does once she has it has mostly been

answered by anecdote. Most of the evidence is about the room. Some is about the people in it. The two are easy to confuse, and the c-factor is where the confusion began.

One number for a room

Here is what Woolley, with Thomas Malone and three colleagues, did. They put small groups of strangers through a battery of tasks drawn from a standard map of what groups do, so that no single kind of task could carry the result. Then they scored the tables. One factor accounted for 43 per cent of the spread in how the groups did across the tasks, and the cleverest person at each table, and the average cleverness of the people at it, predicted that factor only moderately.

That is the first thing to carry out of this chapter, and it survived everything that came later. A group's performance is not the average of its members.

Three things predicted the factor better than cleverness did. The first was how well the people at the table could read a face, measured by a test in which you are shown a strip of photograph containing only a pair of eyes and asked which of the words offered describes what the person is feeling. The second was turn-taking: the more unevenly the talk was spread, the lower the table scored. The third was the proportion of women. And the authors said in the paper that the third was mostly carried by the first. Women on average score higher on the eyes test, so a table with more women had, on average, more capacity to read a face, and that was the thing that mattered. The paper ran in *Science*. The last of the three findings ran everywhere else.

That link needs its layer, so here it is. The eyes-test advantage is one of the better-replicated small differences in the field, found again in dozens of countries, at every age and in almost every language tested. It is a score on a test, so I file it under behaviour: real at the level of a population, with an overlap so large that a great many men sit above the average woman at it. The ability is real, it is not evenly distributed between the sexes, and it is nowhere near a rule about the man or the woman sitting next to you.

The finding spread quickly, as findings do when they tell a room what it wants to hear. A second study found the eyes-test link again in groups

that worked only through text on a screen, which suggested that whatever the test measures is not just the reading of faces but something more general about reading people. A meta-analysis from the same research cluster as Woolley's pooled the studies that had followed and found the structure of c intact. By then c had been in the slide decks for years. More women, smarter teams. It was, in the Preface's terms, the second half of the question answered in one line, which should have been the first warning.

The second look

Psychometricians are the people you send for when a finding has been enjoyed too much. In 2017 Marcus Credé and Garett Howardson published a paper in the *Journal of Applied Psychology* whose title was the warning: a second look at the evidence for the collective intelligence construct. They took six of the published samples on which c had been built and reran the sums.

What they found explained at most a third of the spread in the task scores, well short of the first paper's figure and of what a factor worth the name has to reach. The standard test for whether a single dominant factor is hiding in data of that shape came back just under the line. And when they asked the question the factor was for, whether a table's c predicted how that table did on the next task it was given, the answer was: weakly. They pointed out that the covariance between tasks that did exist had two cheap explanations. Groups that did not try hard did badly at everything, which looks like low c and is low effort. And the data were nested in ways the original analyses had not fully handled. They cautioned, in the paper's last pages, against firms abandoning the selection of able individuals in favour of a group metric that had not yet earned it.

So here is the fall, said once. The factor fell. The claim that a group has one intelligence the way a person has, general and measurable, did not survive its second look, and the headline that had been hung on it, more women, smarter teams, fell with it. What makes a table more than its members is not the number of women at it.

What Credé and Howardson left standing, and what the field now broadly accepts, is the finding underneath. Tables where a few people dominate the talk do worse. Tables whose people can read each other do

better. The mechanism partly survived. The headline did not. A famous finding can fall without its useful part falling with it, and this one did.

And the useful part belongs to nobody's sex. The turn can be written into the rules of the room and handed to anyone. The reading of faces is a score on a test that plenty of men beat the average woman at. What helps a table is how the turns are taken and how the faces are read, and neither comes with a sex attached. Both sides are still publishing, and the meta-analysis from Woolley's cluster is the proponents' reply. I give you the dispute as a dispute, which is what the title of this chapter means: the factor fell under scrutiny, and the scrutiny is not finished.

What the field studies say is messier than either side's headline. The best causal evidence that a balanced team does better at something real comes from a study that put business students into teams and had each team run a real company for a term, selling shares, electing officers, moving product. The sex ratio of each team was assigned by the researchers. The balanced teams outsold and outearned the male-dominated ones, and profit rose with the share of women up to the midpoint and then flattened, with too few heavily female teams to say what happened beyond it. The authors measured conflict, friendship, decision process, atmosphere, mutual monitoring and learning, and none of it explained the result. It comes with no mechanism attached.

Now the other column. A business simulation played in small teams by undergraduates and MBA students found the all-women teams outperformed by every other combination. The reason was in the decisions: the all-women teams priced less aggressively, put less into research and development and more into what the simulation called social sustainability, in a game that paid for aggression. That is not a finding about who is good at business. It is a finding that a game built to reward one end of a distribution rewards the teams drawn from that end, which is the kind of sentence that ought to be printed on the box.

And in 2024 Hardt, Mayer and Rincke put people into small teams and had them solve a sex-neutral puzzle in an audio-only chat room, no faces, no eyes to read. All-male teams talked more about the task and did better than both mixed and all-female teams. The striking number is inside the mixed teams. Men in a mixed room talked more than men in an all-male room, women in a mixed room talked less than women in an all-female room, and the men ended up saying almost 70 per cent more words

than the women. Sort by ability and the ablest women spoke less than the least able men. Take the eyes out of the room and the thing Woolley found, the even turn, goes missing. That is the boundary condition on the whole collective-intelligence story, and it is a finding about a medium, not about a sex.

Do women shy away?

Chapter 4 gave you the headline that sits under all of this. In 2007 Muriel Niederle and Lise Vesterlund sat men and women down to add up numbers for a few minutes, found them equal at it, and then let each choose to be paid by the piece or by tournament. About twice as many men as women chose the tournament. The paper's title asked two questions, do women shy away from competition, do men compete too much, and the data answered yes to both. Men who were bad at the task entered too often and lost money. Women who were good at it stayed out and left money on the table. The sentence to carry is the first one, because it is the one the folklore drops.

What happened to the gap under scrutiny is the pattern of this chapter again. It replicated, smaller. Pool the dozens of papers that followed and the gap is still there, largest when the task is arithmetic and the subjects are students, smaller with verbal tasks, other ages and team competition. It travelled out of the laboratory, into the school choices of Dutch teenagers and into how the sexes carry a loss into the next round: after losing, women cut their effort whatever the prize, and men only when the prize was large. That is the behavioural layer, and it is statistical. The two distributions overlap by most of their width, and pick a random pair and the woman is the more competitive one often enough that you should not bet against her.

Then came the measurement people. Three economists argued in 2019 that the preference Niederle and Vesterlund had found, after allowing for risk attitude and confidence, was an artefact of measuring risk attitude and confidence badly. A control measured with error shrinks toward zero and leaves its residue to be picked up by whatever is left in the regression, which was the variable for sex. When they reran the design on a cohort of Caltech students with better measurements, the sex gap in entry was en-

tirely accounted for by risk attitude and overconfidence. There was no separate shyness. There were women who were on average more risk-averse and more accurately confident than men, and that was enough. A pure replication the year after got nothing at all. Whether the gap is a preference or a symptom is open, and it is a different kind of open from the c-factor's. Nobody disputes that fewer women enter. They dispute what the entry is made of.

The hormone that was supposed to be underneath it has been looked for and not found, and chapter 3 owns that story, so one paragraph. The best trial, pre-registered and large, gave men a dose of testosterone up the nose or a placebo and found no effect on competition, risk or anything else it measured, and an earlier trial had found the same in postmenopausal women. Pooled across every study that has tried, the correlation between testosterone and risk-taking sits on zero. The subtler idea, that testosterone drives status-seeking only when cortisol is low, the dual-hormone hypothesis, got its own meta-analysis in 2019 and came back with an interaction too small to use. It fell, and when chapter 11 meets it again at the scale of the boardroom, it meets it as fallen. For the table, the point is this: whatever makes the extra men enter, a dose of the molecule does not move it.

Culture, the third layer, is where the gap does the thing averages are not supposed to do. It changes sign. Take the same choice to the Khasi of Meghalaya in north-east India, where inheritance and clan membership run through the mother and the youngest daughter takes the family wealth, and the women compete more than the men. Take it to the Maasai of Tanzania, where women hold little property, and the men compete more than the women. The task was tossing a ball into a bucket, and the Khasi women competed more than the Maasai men. Among the matrilineal Mosuo of Yunnan the girls started out more willing to take risks than the boys and, after years in schools shared with the Han majority, were more risk-averse than the boys. Culture did not just set the gap. It reset it inside a childhood. And among Norwegians whose parents came from dozens of countries, raised under the same schools and the same labour laws, the gap is smaller the more equal the parents' country of origin, from both sides: the men from equal countries compete less, the women more.

Change the form

So the gap is real, replicable, smaller than the first paper, possibly made of risk and confidence rather than of shyness, moved by no hormone that has been tried, and reversible by kinship. What can a team do with that? The answer the economists found is almost insultingly practical. Change the form.

In 2015 Flory, Leibbrandt and List placed a real job advertisement in front of thousands of job seekers and varied one thing. In one version the pay was competitive if you ticked the box to make it so. In the other the pay was competitive unless you ticked the box to make it not. With the competition as the thing you had to choose, the men applied far more often than the women, a gap of nearly 26 points, about the size of Niederle's. With the competition as the default you could leave, the gap was a single point. It did not shrink. It went. Nobody's preference changed. The form did what a generation of advice to women had not, by not asking. And a quota that guaranteed women an equal share of the winning places in a tournament pulled in capable women who had been staying out, and did not lower the standard a winner had to meet, because the women it drew in were good.

The rest of the architecture of a workplace carries the same lesson. Show evaluators the files of candidates who have asked for more salary and they penalise the women who asked far more than the men who asked, and put the same people in the candidate's chair and the women decline to negotiate with a male evaluator, not from a lack of nerve but because they have priced the penalty correctly. Have thousands of adults and schoolchildren do a maths task and rate their own performance, and with actual performance held constant the women rate themselves about a quarter lower than the men, in private as well as in public, and already in children at school, and employers who see the self-ratings hire the women less and pay them less. Ask who is willing to do the task that helps the group and promotes no one, and women volunteer half again as often as men, managers of both sexes ask women to do it more often, and women say yes more often when asked. Everyone in the room expected the woman to do the unrewarded thing, including the woman.

Every one of those is a finding about a form, a default, a chair or an expectation, with a behavioural difference underneath it that is real on

average and overlapping. The three-layer sentence for a team is this: the capacity to read a face and the appetite for a tournament are measurable and differ by sex at the level of the population. What they do at the table is statistical. What the table asks of them is ours.

Scale it up once and the same shape holds. Norway's boardroom quota was found, in a paper that went round the world, to have cut the value of the country's firms. A decade later the sums were redone, the events turned out to have been dated wrongly, and the effect was indistinguishable from zero. Chapter 11 has the boardroom. I mention it here because its headline fell exactly as the c-factor's did, and for the same reason.

One more study, which is not about people. Train a minority of golden shiners to swim toward a yellow target and a majority toward a blue one, and the minority steers the school, because it wants it more. Add a few fish with no training and no preference, and the school goes where the majority wants. The uninformed restore the majority. A fish tank is not a boardroom, though some boardrooms have been compared to one unkindly. The lesson survives the move. A body's decisions are shaped by who is in it and by how many in it hold no brief, and the second is a design choice.

Here is the turn. Everything in this chapter that fell, fell because somebody had read a composition as a cause: more women, more c. More women on the board, less value. More testosterone, more entry. Everything that held was about the room: the turn, the medium, the default on the form, the chair you ask from, the task nobody is paid for, the fish with no opinion. The accent is in the people, and it is real. The factor is in the room, not in the sexes, and the room is the thing you can change by Monday.

THE STRONGEST CASE AGAINST

The chapter has argued that the sex of the people at the table matters less than the rules of the table. Here is the case that it matters more than I have let it, and its best voice is Thomas Buser's.

Buser, with Niederle and a colleague, measured the competitiveness of Dutch fifteen-year-olds in a classroom and then followed them into the tracks the Dutch school system offers. The laboratory measure accounted

for about a quarter of the sex gap in who chose the prestigious, mathematical tracks, with ability, risk and confidence held constant, before any employer or any form had seen them. The measurement critique explains the entry gap, it does not remove it: whichever word you put on it, fewer women enter. The way the sexes carry a defeat into the next round is a difference no form fixes. The self-rating gap was there when nobody was watching, and it was there in children.

Then the teams themselves. The all-women teams in the business game scored below every other combination because of the decisions they took, not because of a form. Hardt's all-male teams beat mixed and all-female teams on a puzzle with no sex in it. The balanced teams that ran real companies did better, but the curve flattened at the midpoint and the authors could find no mechanism in anything they measured. The Khasi are one society, the Mosuo reversal was erased by school, and Flory's opt-out form equalised who applied and said nothing about who performed or who stayed. And Woolley's three correlates were never refuted as behaviour: Credé and Howardson left standing the finding that tables with higher eyes-test scores and more even turns do better, and the eyes-test advantage now rests on one of the largest samples in the field. On that reading the women at the table were carrying something real, the c-factor was only a bad name for it, and a chapter that files the whole thing under the room has mistaken the container for the contents.

I think that reading is about half right. What neither reading can do yet is name the mechanism. Nobody has found the mediator in the company teams, the hormone in the trials, or the preference once the measurement is fixed. Until someone does, the room is the only part of this a team can act on.

IN PRACTICE

If your team is small and one of you is the only woman or the only man, then the research on tokens applies: the solo woman in a mostly male group talks less and carries less weight. Put the turn-taking in the structure, not in the goodwill: a round in which each person speaks before anyone speaks twice, a written first pass before the spoken one. The failure mode is the round that becomes a ritual while the real decision happens afterwards in the corridor, which is where it was happening before.

If your team works by voice alone, by phone or in a chat with the cameras off, then expect the Hardt pattern, men speaking more than they do in a room of men and women less than they do in a room of women, and expect the able women to be the quietest. Turn the cameras on, or move the first round into text, where the groups in the one study that tried it still read each other. The failure mode is that it is one study against Hardt's, and the medium may decide more than any rule you add to it.

If you fill the hard assignment or the promotion by asking people to put themselves forward, then you are running Niederle's tournament with the opt-in form, and in Flory's field trial that form lost most of the women before anyone had been assessed. Make the competitive track the default and let people leave it. The failure mode is that opt-out equalises entry, not outcome: on the measurement reading the women who stayed out were right about their own risk, and some who are now in tend to lose, as the low-ability men in Niederle's data did.

If pay in your team is set by negotiation, then the penalty for asking falls on whoever is seen as asking against type, and in the data that was women. Publish the bands. The failure mode is that a band removes the ask and with it the one lever the strongest negotiator had, and nobody has measured what that costs a team.

If the tasks that keep the team running but promote no one are filled by asking who is willing, then both sexes expect a woman to say yes, the manager asks her more often, and she says yes more often. Rotate the task by list or by lot. The failure mode is the list that quietly exempts the person whose time is deemed too valuable, which restores the old allocation under a new name, and nobody has shown that rotation improves a team's output, only its fairness.

If the team is deciding something with a real downside, then there is one honest reason to want it mixed, and it is not that women are careful. It is the fish. A body whose members do not all want the same thing hard is harder to steer off a cliff, and a body in which some members hold no brief steadies toward its majority. The failure mode is the central bank, where more women on the committee went with higher interest rates for the same expected inflation. The women were the hawks, which was not in anybody's pamphlet, and the authors' guess is that the profession's filters let only the hard ones through. If your door selects for one temperament, mixing the sexes behind it tends to change nothing you can measure.

If you were hoping to make the team smarter by adding women, then do not promise it to anyone. The factor that underwrote the promise did not survive its second look. What you can say is that a team in which everyone speaks, and in which people can read each other, tends to do better across tasks, and that the share of women in a room has, on average, moved both of those in the laboratory. On average. The man across the table may be the one who reads the room best. Look at him, not at the ratio.

And if you are the one woman at the table, or the one man, then the room is reading you as a category before it has heard you, and the quietest version of you is the one it is likely to get. The one thing that reliably changed that in the data was a rule that made the turn yours before you had to take it. Ask for the rule. The failure mode is the room that gives you the turn and takes the decision elsewhere, and the counter-evidence is the company teams: where nobody was the only one, no rule was needed, because the numbers did the work.

The studies, figures and open questions behind this chapter are on the website.

PART THREE

Emancipation 2.0

CHAPTER II

The Abbess and the Monks

A woman is chairing a meeting on the fourth floor of an office building and her face has started to burn. It began at the collarbone a moment ago and is climbing. Everyone else in the room is comfortable. She does not touch her collar or reach for the water, because she has learned that the one thing harder to bear than the heat is being seen to notice it. Across the table a man of my generation is in the fifty-sixth hour of his working week, and he does not notice that either. He is proud of it, in the quiet way men of my generation are proud of things we were taught not to mention.

Neither body is doing anything unusual. Hers is finishing its reproductive years at the usual age. His is running a week that, in the largest pooled studies of working hours, goes with markedly more strokes than an ordinary one. The room was built for neither of them. It was built for a body that is never cyclical, never pregnant, never tired, and has somebody at home. That body does not exist. It has run most organisations for about two centuries.

Whose body was the rule written for? Ask it of every organisation you have belonged to, as this chapter asks it of every one in the record.

Nine centuries ago a wandering preacher wrote a rule for a house in which the men took their orders from a woman. Robert of Arbrissel founded the abbey of Fontevraud, in France, in 1101: a house of monks and a house of nuns, kept strictly apart, both answering to the abbess, to whom the monks owed the obedience a son owes his mother. It was still standing when the French Revolution closed it in 1790.

That house is the chapter's picture. Not because a woman in charge is

the point, which it is not. Because its rule treated the two questions every organisation has to answer, who does what and who answers to whom, as two separate questions. Most organisations answer them as one. So, probably, does yours.

What the abbess does differently

The organisation chart was supposed to be written in the measurable layer, by hormones. It was not. The idea that testosterone picks the leader when the stress hormone is low had its decade and fell, as chapter 10 told, and what survived the wreckage points the other way: the job sets the hormones far more reliably than the hormones pick the job. That is the Endocrine Accent from the other side: the firm is one of the things that sets the lilt.

Now the behaviour, which is statistical, and here the two bells are very nearly one bell.

Alice Eagly spent a career testing the stereotype everyone knew, that women lead through relationships and men through the task. In 1990 she and a colleague pooled every study of leadership style they could find and looked at what happened in real organisations rather than in laboratories. The task difference came to nothing. The relationship difference came to nearly nothing. The one difference that survived was a style of decision, women leaning democratic and men autocratic, by a margin so small that the two curves lie almost on top of each other. The critics said, fairly, that too many of the studies were leaders rating themselves.

The largest study of rated effectiveness found no difference when all the ratings were pooled. Then the authors split the raters. Men rated themselves higher as leaders than women rated themselves. Everyone else, bosses, peers and the people underneath, rated the women slightly higher. The authors' reading is a filter: a woman has to clear a higher bar to get the job, so the women who hold it are on average the stronger ones. Whatever the explanation, a difference that appears when you ask the leader and reverses when you ask the led is not a difference in leading.

Where the sexes do differ is in who gets to lead in the first place, and that difference lives in the room. The best study of who emerges as leader in a group that starts without one found men, still, though the gap has

been closing for decades. The mechanism is talk. Assertive people speak more, the group hands the chair to whoever has been speaking, and assertiveness runs higher in men on average, with the usual wide overlap. The counter-mechanism is in the same data. Once a woman has the chair, the ability to read the room, which runs higher in women on average, predicts that the group rates her leadership as the good kind. Emergence rewards the accent of agency. Performance rewards the other one. An organisation that promotes by emergence has chosen, without noticing, which it is paying for.

Then the cultural layer, which Eagly also named. Role congruity is her term for the fact that the leader stereotype and the female stereotype point different ways, so a woman who leads the communal way is a weak leader and a woman who leads the agentic way is an unpleasant woman. Show raters the same scripted autocratic behaviour and they mark it down in a woman and let it pass in a man. And two researchers reading board appointments at Britain's largest listed companies found women appointed to the top disproportionately when the firm was already failing. They called it the glass cliff. Think crisis, think female, and a scapegoat is always useful to have about the place. The cliff is the only chair on offer.

That is the behavioural layer. Nearly nothing in the leading, something in the emerging, a great deal in the watching. The sex of the person at the top changes very little about what the house does.

The rule of the house

Now the history, which is this chapter's one real advantage over the two shelves: they argue about what an organisation built around the difference would look like, and the record has several.

The Anglo-Saxon church built one first. Between about 600 and 900 the double monastery, monks and nuns in one foundation under one abbess, was the mission station of the English conversion, and the abbess was very often a royal widow or a king's daughter. Hild of Whitby is the name that survives. The theological cover was that people who had given up sex, property and will had given up their sex as well. One historian objects that the chroniclers praised the great ascetic women for acting *viriliter*, like a man, so what the cloister offered a woman was not neutrality

but promotion to honorary man. I think she is right, and I think it was still the best deal available in 700.

The East said no. The Second Council of Nicaea in 787 ruled that women ought not to live in monasteries of men, and banned new double houses in the Orthodox world. The West did not follow. In 1101 Robert of Arbrissel built Fontevraud on the model the East had just forbidden, and then improved it. Separate houses, strictly kept. One abbess over both. The Gilbertines at Sempringham in England went further on the small things, where rules live: the same food and drink for the nuns as for the canons. Separate quarters. One table. One rule. And the rule said who answered to whom in so many words, so that nobody had to find out by seeing who spoke most.

Here is the thing to notice, because the rest of the chapter rests on it. Who leads and how the house is run are two questions. Most firms answer only the first, and believe they have answered both.

The guilds answered the same questions the other way. Sheilagh Ogilvie, whose history of the European guilds rests on centuries of their own records, found that guilds were cartels, colluding with the political class to restrict competition, and that one of their standard ways of keeping members' wages up was to keep women out. Her earlier work on Württemberg makes the point that matters for the layers: the guilds had to pass regulation after regulation against women weaving, brewing and trading, and a rule that has to be re-passed every generation is a rule the bodies keep breaking. The court rolls of one Northamptonshire village show brewing in the early fourteenth century as women's work. When hops arrived and beer kept and the trade needed capital, it became a learned art, and learned arts were for men. In 1675, with Colbert behind them, the seamstresses of Paris got a guild of their own, with a monopoly on clothing for women and small children, and by the 1750s it was taking in more apprentices than any other trade in the city. Given the licence, the women ran a cartel exactly as well as the men. Which is not a compliment to either.

Below both, most people lived in a third arrangement with no rule written at all, the working couple of early modern Germany, his portfolio the plough and the burst strength, hers the dairy, the textiles and the market, and neither solvent without the other. Three designs, then, before a single factory. Separate houses under one rule that named the authority.

A cartel that wrote the body into the rule. And a couple that wrote nothing and divided by what each body did best. The chapter is about which of them we rebuilt.

The worker with no body

We rebuilt the cartel, and gave it a factory.

The pre-industrial household was the firm, with subsistence, market production and children in one place, and industrialisation moved the first two out of the house and left the third behind. Women's paid work fell through early industrialisation, as work left the home and a rising male wage bought a wife's absence from the labour market, and rose again only when offices wanted education more than muscle. The English middle class poured an ideology to fit the shape, which two historians found in its wills, sermons and house plans: commerce as male and corrupting, the home as female and the nursery of virtue, a wife's idleness as the mark of a man who had arrived.

Out of that came the body the rule was written for. Sort the welfare states by how hard they built the breadwinner into tax and childcare and Britain and Germany are the strong cases, Sweden and Denmark the weak ones. I was born in 1975 into the strong column. What I want here is the worker the strong column assumed: paid, continuous, available, with the laundry done by someone else. Sociologists call him the ideal worker. Here is what he costs the two bodies that actually turn up.

Her first. A large Dutch survey asked women about their periods and their work. Few had stayed home for them. Most had gone to work in pain and got less done, and over a year that came to about nine working days of output lost without a single absence on the record. The absence the rule can see is small. The presence it cannot see is several times larger, and nobody has yet measured it from the employer's side. Now the laws. Japan has had menstrual leave since 1947, and fewer than one woman in a hundred uses it. South Korea mandates it too, and most women leave it untaken. Jonathan Gruber showed in 1994 that when a state mandates a benefit only one sex uses, employers price it in, and the cost comes back to that sex as lower wages and fewer offers. The leave is a cost on her payslip before she has taken a day of it.

The night shift is the clearest case in the chapter of a design decision landing on a body. Pool the studies of women on shift work and the cycle gets more irregular and more painful, and the menopause comes a little earlier. Fixed nights go with more miscarriages, very long weeks in pregnancy with more preterm births, afternoon and night work with more gestational diabetes. Nobody ordered any of this. Someone wrote a rota.

Then the end of the cycle, which is where the woman at the table started this chapter. Nearly every woman has symptoms through the transition. The Fawcett Society asked British women in 2022 what it had done to their working lives, and one in ten had left a job because of it. Others had cut their hours or turned down a promotion, and a good many said their colleagues had treated it as a joke. The British Standards Institution published a standard for it in 2023, unglamorous and specific: breathable fabrics, less time in heavy protective kit, a fan, a window, shifts moved on request. Against which the researchers who study the menopause at work warn that a firm that files it under medicine files the women under decline, an older man experienced and an older woman symptomatic. Both are right. The window and the register are different things.

Put her side together and it comes to this. The body at work is a scheduling problem, and most firms have spent the whole age of the factory pretending not to have it.

His body now, and his is the one this book's shelves usually leave out. Mika Kivimäki and an international consortium pooled the working cohorts of Europe, the United States and Australia and published the result in *The Lancet* in 2015. Against an ordinary week, a week of fifty-five hours or more went with about a third more strokes, and more heart disease, the risk climbing in steps well before that threshold, after allowing for smoking, drink, inactivity and class. The World Health Organization and the International Labour Organization then did the sums across the world's labour surveys and put the deaths from heart disease and stroke attributable to the long week, in a single year, at about three-quarters of a million. Most of them were men. The two agencies named the long week the occupational risk factor with the largest burden of disease. Not asbestos. Not the mine. The week.

The mine is in there too. In the United States, year after year, the dead at work are almost all men, from a workforce that is nothing like all men, and the ratio has not moved in a generation of safety law. The reason is

sorting, not daring: structural iron, roofing, fishing, logging, construction and the heavy roads are where the men are. And when the work stops, unemployment goes with more suicide, most in the first years and more in men than in women, though a good part of the raw figure is ill health that came first. The provider role is a cultural script. It is written on a body that reads it as a verdict.

The script runs the other way too. A survey of fathers in large American firms found nearly all of them calling flexibility critical and a great many saying they would consider staying home if the money allowed. They do not, because a man who asks for family leave is read as less committed and, specifically, less masculine, and is rated and promoted accordingly: the flexibility stigma. Men given flexibility often work more hours than before, to prove they did not want it. So the ideal worker is kept in place from both sides. She is priced for a body she has. He is penalised for asking to have one.

Notice what the costs have in common. Not one of them came from who sat at the top. They came from the rota, the week, the leave form and the ratings sheet. What changes outcomes is the rule of the house, how the work is scheduled, who may interrupt it and what gets measured, and not the sex of the person at the top.

Separate doors

The last question is the oldest: when the bodies differ this much, do you build one house or two?

The answer depends on which layer the task lives in, and the military is the cleanest test because it runs on all three. A review for RAND in 1996 took apart the word cohesion, the whole argument against mixing: what predicts a unit's performance is task cohesion, the shared commitment to the job, not social cohesion, whether the members like each other.

Then the task crossed into the measurable layer. In 2015 the United States Marine Corps ran a year-long experiment with female volunteers in otherwise closed combat units and measured mixed teams against all-male teams on every task it could time or score. In about two tasks in three the all-male teams came out ahead, and the widest gaps were in the ones that matter most: carrying a casualty, marching under load, clearing

an obstacle. Women were injured more. On the infantry training course most of the women did not graduate, and nearly every man did. Israel's mixed battalion, Caracal, gives the other half: its women had more stress fractures than its men and dropped out less, because the women who chose it had chosen it harder. Sport draws the same line, a male advantage from puberty that is widest where the task is strength, and a year of testosterone suppression removes only a sliver of it. Where the task is the body, two doors. Where the task is anything else, the evidence for two doors is thin. Chapter 9 has the single-sex schooling result and its rounding error, and one study of pupils found the one reliable product of a single-sex school was more anxiety around the other sex.

But separate is not the same as subordinate, and the record's most durable arrangements kept the doors and shared the top: chapter 2's Haudenosaunee, whose male chiefs were nominated and deposed by the Clan Mothers. And there is a number for the share at which a minority starts to count. Bina Agarwal studied the community forest committees of Gujarat and Nepal at the turn of this century and found the all-women committees of Nepal handed the smaller forests and getting more regeneration out of them, and found women's voice on a mixed committee beginning at roughly a quarter to a third of the seats. Below that, a woman on a committee is a woman in a room.

The kibbutz is the control. From 1910 the movement set out to abolish the sexual division of labour outright, with communal kitchens and children's houses, and by the third generation the fields and the building sites were about nine in ten men, the children's houses and the kitchens about the same proportion women, and a man far likelier than a woman to hold a senior post. The study that counted called it biology. Its critics say the women were defined out of the fields by managers who called only heavy work productive, or that the return to the family was a revolt against an equality that meant living like men. Either way the organisation that wrote no rule got the old one back.

Which is the last thing to carry. The house that answered both questions was designed. Nobody has yet grown one by leaving it alone.

Here is the tool, then. Every organisation answers two questions, and the record's failures answered them as one. Who does what is a question for the bodies and the bells: test the task and never the sex, and expect the sorting to be lopsided at the extremes and overlapping everywhere else.

Who answers to whom is not a question for the bodies at all. It is a question for the rule, and if the rule does not answer it the room will, by whoever speaks most, and the room has a known accent. Robert of Arbrissel wrote the second answer down. We mostly let it emerge.

THE STRONGEST CASE AGAINST

The chapter has argued for writing the organisation around the bodies it actually employs. Here is the best evidence that writing the body in is exactly the mistake.

Every time a rule has named a female body, the owner of that body has paid for the naming. Gruber's result from the American states is a measurement, not a prediction: mandated maternity benefits came back to women of childbearing age as lower wages, nearly in full, because employers treated the mandate as a tax on hiring them. The same has since been found across the rich world for every sex-specific leave that can be measured, hardest on women in small firms, precarious contracts and male trades. Japan and Korea wrote menstrual leave into law and the women leave it alone. A small majority of young women, asked, say they would like an explicit policy, and the economists' answer is that its cost is in the offers they never see. On this reading the window and the fan are harmless because nobody has to be named to use them, and everything in this chapter that names a body is a tax on it.

The second argument is that the body was never the reason. A reading of the 1833 Factory Inquiry found women barred from the overseer's job, the best paid in the mill, long after steam had taken the strength out of it, which means strength was a word for a door, not a measurement of one. And the Marine Corps numbers are numbers about volunteers held to a standard built for men's bodies, in units closed to women the year before. The third argument comes from the other shelf. The kibbutz says that when an organisation refuses to write the body in, the body writes itself in anyway, and on that reading the abbess is not a design but a tail of the distribution with a crown on, which nobody else can live under.

I think the first argument is right about labels and wrong about design, and the window is the proof. I think the second is right about the mill and does not reach the casualty carry. And I think the third describes exactly

what happens when the two questions are answered as one. But the first is the one I cannot fully answer, because the firm that writes nothing down and the firm that names the body both have a bill, and nobody has yet said which is larger.

IN PRACTICE

What follows is for a team, a firm or a public body, in conditionals, with the way each one fails.

If your organisation promotes the people who emerge, which means the people who spoke most in the meeting where no one had the chair, then the emergence data say you are selecting for the agentic accent and paying for it in the ratings afterwards. Decide the chair before the meeting and write the criteria for the job before the candidates are in the room. The failure mode is that the written criteria turn out to be the leader stereotype in a numbered list, which is still masculine, so read them aloud and ask which sex they describe.

If you are about to appoint a woman to a post that is already in trouble, then the boardroom data say you are about to do the usual thing, and the usual thing ends with the post and the woman both blamed. Give her the runway, the budget and the margin for error that the last man had. The failure mode is that nobody can tell, from inside the crisis, whether they are being brave or conventional, so look at who got the last few such posts in your own records.

If the work runs on long weeks and you are the senior man who sets the tone, then the cohorts say the week is the risk factor, not the stress you can feel, and the stigma data say the junior men are watching your hours and not your policy. Cut your own first, visibly. The failure mode is that a man given flexibility tends to work more to prove he did not want it, so the policy that lands is the one the most senior man is seen to use.

If you want to accommodate the cycle, pregnancy or the menopause, then the labelled leave of Japan, Korea and the rest says the label is a tax and the uptake is close to nothing, while the window, the fan, the shift moved on request get used, because nobody has to be named to use them. Build the second kind and make it available to everyone, which means the man with the migraine too. The failure mode is the register: the moment

a firm keeps a list of its menopausal women it has a list, and the warning about who is experienced and who is symptomatic applies.

If you run a rota with fixed nights and a pregnant woman is on it, then the pooled studies say a day role for the duration tends to lower a real risk, and the economics say that offering it to her alone tends to price her. Offer the move to any worker with a stated medical reason, at the same pay, and do not ask which reason. The failure mode is that a quiet offer is a quiet penalty, so the offer has to be the rule and not the exception.

If the task is the body, a casualty carry, a load over distance, a contact sport, then the Marine Corps and Caracal data say the sex gap at the extreme is large and does not wash out with training or with hormones, so test the task, at the standard the task needs, and let whoever passes pass. The failure mode cuts both ways. A standard set above the task's real need is the overseer's door with a stopwatch, and a standard quietly lowered for one sex is a graduation rate waiting to happen in the field.

If your organisation has let its division of labour sort itself and finds, like the kibbutz, that it has sorted along the old lines, then the lesson is not that the lines were right. It is that the one question you forgot to answer in writing was who answers to whom. Answer that one on paper, as the abbey did, and let the other one sort. The failure mode is the oldest in the chapter: separate becomes subordinate the moment the authority is left to emerge, and the Second Council of Nicaea is still there to tell you which way it emerges.

And if none of this fits your organisation, then the overlap is doing its work. Every finding above is an average, and the woman at the table on the fourth floor is one person, in one room, with one rule, which somebody, at some point, chose.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 12

The Bus Route

Stand at a bus stop at ten past nine on a weekday morning, in any European city you like, and watch who gets on. The commuters have gone. The bus that comes now is half empty, and the people who board it are carrying things: a pushchair with a toddler asleep in it and shopping hooked on the handles, a man with a child's scooter under one arm and the child under the other. The woman with the pushchair rides a few stops to the clinic, later a few more to the supermarket, then home on a transfer that has expired, so she pays again. The man with the scooter has dropped one child at nursery and is on his way to collect another from a grandmother. By lunchtime each has made a handful of short trips that add up to a loop. The bus route under them is a line.

That line was drawn for someone. It was drawn for a passenger who leaves home at ten to eight, rides to the centre, works, and rides back at six, and it is a very good line for him. The timetable thickens at his hours and thins at everyone else's. The fare charges him once each way. The map is a star with the office district at its heart and the suburbs at the points, and the route from one suburb to the next, nursery to grandmother, goes through the centre or does not exist. He was, for most of the twentieth century, a man, and the people riding the loop at ten past nine were women. I rode routes like his for years without once noticing who else was on the bus, which a man born in 1975 can admit without much credit.

This chapter is about the largest thing human beings build and the fact that it was built with one passenger in mind. It is also the chapter where I know least: much of what I know about the city that did it differently is

that city's own account of itself.

Ride the loop with them. Watch where the bus does not go.

Lines and loops

Transport surveys used to count two kinds of trip as real: going to work and going to school. Everything else went into a bin marked other, or escorting, or shopping, or leisure, and a bin is not a category anyone plans a bus route for. Inés Sánchez de Madariaga, an urban planner, gave the bin a name in 2009 and a framework in 2013. She called it the mobility of care: all the travel done to keep a household running and its dependants alive, which is to say the nursery run, the clinic, the chemist, the shop, the grandmother, in whatever order the day demands.

Once the bin had a name, cities started counting what was in it. The best count I know is Montréal's. A regional travel survey, gone through again by a team of researchers, found that care accounted for 28 per cent of all trips in the city: nearly a third of women's daily trips, and a quarter of men's. Washington's fare cards tell the same story: women are most of the city's transit riders, and more of them still on the off-peak, chained trips. Two Latin American cities found a smaller share and the same shape. A large slice of urban travel is care, and almost none of it was ever planned for.

Cities were drawn for one kind of trip, home to work and back, and most trips are not that kind.

Read Montréal's numbers for the overlap before you read them for the gap. A quarter of men's trips are care trips. The loop is not a female shape. It is the shape of whoever is doing the care, and a father who is the primary carer makes exactly the same chained, off-peak, short-hop pattern as a mother. The bus does not check your chromosomes. It checks whether you are pushing something.

So the people riding sideways are mostly women, and the children and the old they are carrying or collecting, and that is a fact about the timetable of care, not a fact about the sexes. The loop is behaviour, and behaviour is statistical: in every city measured, more of women's travel is care than men's, by a margin of a few points, and the margin is not a rule for the woman with the pushchair or the man with the scooter. What puts her

on the loop more often than him is culture, the ledger of minutes chapter 7 went through, and under the ledger sits a physiology, the gestation and lactation that chapter 8 priced, which starts the sorting before anyone has drawn a map. The route itself is none of these. It is a decision.

The decision has a price, paid at the fare box and on the map. Most systems charge by the boarding, not by the hour, so a morning of short hops costs several times a single long ride, and the person paying several times is the carer. And a hub-and-spoke network does not run from one suburb to the next, which is where the nursery and the grandmother are. The carer who can afford a car buys one. The carer who cannot is left with what the literature calls time poverty, the quiet tax on a loop that was never counted.

Who owns public space

In 1991 a planner in the Vienna city administration named Eva Kail put on a photography exhibition. Its title, in translation, was *Who Owns Public Space: Women's Everyday Life in the City*, and what it showed was the loop: women photographed through their days, in the stairwell with a pram, at the bus stop, in the park with the children, in the queue. It set off a public argument, which was the point. The year after, the city set up a Women's Office, and in 1998 a planning unit called the Co-ordination Office for Planning and Construction Geared to the Requirements of Daily Life, a title only a city administration could love and a function only a city administration could deliver.

In 1999 the city surveyed how its residents moved, and the survey drew the picture this chapter opened on. Men's trips were lines, home to work and back. Women's were loops, varied, local and tied to the small infrastructure of the neighbourhood: the nursery, the shop, the doctor, the stop. I know the survey only at second hand, and I say so because it is the origin of this chapter's governing image.

What Vienna built on that picture was not a theory of women. It was a procedure. Who sits at the planning table, where the money goes, what gets built, what the law allows: the city made each of those a question every new development had to answer, and in 2013 Kail and a colleague wrote the questions into a manual. Gender mainstreaming is the term of

art, and for once the term is the right one. The city sorted by gender, by who did what, and measured sex, the body, only where the body makes its own demands, which we come to at the toilet door.

The projects that carry the chapter are a housing estate and a park.

The housing first. In the late 1990s the architect Franziska Ullmann and others designed Frauen-Werk-Stadt, Women-Work-City: an estate of subsidised flats with a kindergarten, a medical centre and shops built into the site, so that the loop could be walked rather than ridden. The flats were arranged around grassed courtyards so that a child playing could be seen from the kitchen window. Stairs and entrances were widened and lit so that there were no dark corners and a pram could get up them. You recognise the shape from chapter 2. A foraging camp is a clearing where the children are visible from where the adults work, and where any adult who looks up is, for that moment, on duty. The courtyard is the Campfire Contract poured in concrete, and the kitchen window is the fire.

Notice who the courtyard is for. Nothing in it asks who is pushing the pram. The father with the scooter walks the same short path to the kindergarten, the grandmother collects from the same door, and the man who never pushes anything climbs a lit stair with no corners in it. When Vienna redrew a piece of the city for the loop, the trip got shorter for everyone who walked it, and nothing in the drawing said the walker had to be a woman.

The scheme was scaled up at Aspern Seestadt, a new district on the city's edge the size of a town, where a developer that did not provide pram storage, communal rooms and a safe lighting plan was denied the public subsidy and the contract. The residents, polled, say they like living there, and the poll is the developer's.

Now the park. Kail and a landscape architect colleague had watched Vienna's parks and found that girls disappear from them after about the age of nine, while boys of the same age become the dominant users. The reason, in the workshops, was the cage: the fenced football or basketball pitch with its single gate, which a group of older boys could hold as territory and which a girl could enter only by walking into a space she could not easily leave. The girls said so: they did not want to be shut in with boys they did not know, and when they tried to share the pitch they were shouted at or worse. At Einsiedlerpark in the fifth district the city opened the fences on several sides, widened the paths, improved the lighting and

spread the things the girls had asked for along the edges, where you could reach them without crossing the pitch: hammocks, seating platforms, a volleyball net, badminton courts. Girls came back, and stayed longer, and from 2005 the lesson became a binding guideline for every park the city designs. The age and the outcome are Vienna's account of Vienna, and nobody outside has yet checked them.

Notice what the park did not do. It did not decide that girls are timid or that boys are territorial. It observed who held the space and who left it, and it changed the fence. That is the whole of the Vienna method, and it is why a chapter on cities can live, a little uneasily, in a book about hormones. The accent, whatever it is, is heard in the park. The fence is ours.

The queue and the dark

Now the body, because there is one place where the city measures sex and not gender, and it is the lavatory.

Clara Greed, writing in the *Town Planning Review*, has spent a career insisting that public toilets are a planning question, not a plumbing one, because they decide who can be out of the house for how long. Building codes have historically answered the question with spatial equality: the same floor area for men and for women. The result is a queue outside one door and none outside the other, and when the queue was finally put into a mathematical model, the model showed why. A woman takes, on average, about a minute longer in there than a man. That is a measurable fact about bodies and their circumstances, and the circumstances are mixed: menstruation and pregnancy belong to the body, the clothing to the culture, and the child or elderly relative who has to be helped is the loop again, following the carer into the cubicle. On top of the time there is the hardware. A men's room of the same floor area holds more fixtures, because urinals pack tighter than cubicles. And queues do not grow in proportion to the difference that feeds them. Near capacity they grow much faster, so a longer service time, plus fewer fixtures, produces a queue out of the door on one side and an empty room on the other.

Every woman who has waited through the interval while the men's door swung freely already knew the finding. It took a queuing model to make it a planning issue.

The model offers a choice of remedies. One is to abandon equal floor space and give the women's side the fixtures the arithmetic asks for, in a ratio set by the queue rather than the architect. The other is to take the sign off the door. Run the model at different intensities of use and a shared room of enclosed cubicles always shortens women's waits, and shortens men's too, or lengthens them only slightly, depending on whether the urinals survive. The public argument about shared toilets has been about trans people. The queuing result is about throughput for everyone. Whether people accept the shared room varies a great deal from one culture to the next, so the model is not the building.

Then the night, where the evidence turns against the chapter, so I want to give it room.

Women say they are more afraid in cities after dark than men do, consistently, and the fear is not an opinion. It is a restriction: it alters which routes are taken, which transport is used and whether a person goes out at all, and it does this whether or not the objective risk matches it. Every city's answer is more light, and whether light reduces crime is one of the longest-running arguments in criminology.

The case for the lamp is a systematic review of the British and American evaluations, published in 2008, which found that improved street lighting cut crime in the lit areas by about a fifth against the unlit controls. Then the odd detail. Night-time crime did not fall more than daytime crime. The reviewers took this to mean the mechanism was not surveillance but something they called community pride: the city visibly spending money on a street, and the street's residents answering with more order of their own. The lamp as a signal, not a floodlight.

The case against starts small and ends large. Paul Marchant, a statistician, spent the best part of a decade arguing that the review's arithmetic was wrong at the root. Crime is clumpy, not evenly scattered, so the usual error bars are too narrow. And lights get installed where crime has just spiked, so a fall afterwards is what you would see anyway as the spike subsides. On that reading the fifth is an artefact of where councils put lamps. Then, in 2015, a team went through the lighting records of dozens of local authorities across England and Wales, over the years in which councils had been dimming and switching off lights to save money, and for the offences most likely to happen at night, burglary, robbery, violence and sexual assault, found no association between the lighting changes and crime.

One way to square the findings is scale: a lamp in one bad alley may shift what happens in that alley, while the lighting level of a whole county does not move its crime rate. I find that plausible and I notice that it is also convenient. What survives either way is the fear, and the planner's move is to separate the questions. Light is a contested tool against crime. It is an uncontested tool against fear, and fear is what keeps a woman off the last bus. The fear is cultural in origin, unevenly spread between the sexes, and its effect on who is at the stop at eleven at night is as real as the timetable.

A city for the loop

A city does one more thing besides moving people, housing them and lighting them. It introduces them. For most of human history mates were brokered by the people around you, family, friends, the congregation, the street, who did the vetting by knowing everyone, and the camp that chapter 2 described worked the same way. The city replaced the camp with the neighbourhood, and the neighbourhood is now being replaced by a screen: around 2013, in America, meeting online overtook meeting through friends. And when psychologists gave the participants at speed-dating events a long questionnaire on every trait and preference the literature says matters, and set a machine-learning model to predict who would want whom, the model could predict a little of how much a person liked people in general, a little of how much people in general liked them, and nothing at all of the desire between two particular people. The thing the apps filter on is the thing that does not predict the only outcome anyone wants. The thing that does produce it, two people in the same place by accident, is what a bus stop, a park and a courtyard are for.

Before the chapter sounds like it is about women, the other column. In Germany one person in five says they are lonely at least sometimes, and the people most at risk are below the median income, male, or from a migrant background, young and middle-aged more than old. In America the share of young men saying they had had no sex at all in the past year rose steeply between the turn of the century and the late 2010s, to nearly a third, and the rise fell hardest on men with low pay, part-time work or no work. Money does not buy the old arrangement back: when fracking raised the wages of men without a degree in particular districts, births rose

and marriages did not. The loneliest people in the city are men on low incomes, which is not what a chapter about building cities for women set out to find, and I would rather say it than tidy it away. The places where they used to meet someone have been replaced by something that cannot do the job.

So what would a city built for the Campfire Contract look like? Not, I think, like a city built for women. A city that fixes women as the carers because they are, statistically, the carers, is the naturalistic fallacy with a planning permit. The aim is to take the friction out of the loop for whoever is walking it, so that the question of who walks it can be decided at the kitchen table and not by the timetable.

The nearest existing proposal is the fifteen-minute city, the neighbourhood in which work, shops, the doctor and the school are a short walk or ride away, and it fits the loop well. Leslie Kern, a feminist urban geographer, warned in *Feminist City* in 2020 what happens if the market builds it: the grocer, the nursery and the café go upmarket, the low-income women who do most of the city's paid care are priced out to the edge, and then commute back in to do it. Proximity as a luxury good undoes itself. Her conditions are social housing, rules against displacement and care provided collectively, which is to say Vienna's conditions, and Vienna has them because it funds housing.

One last thing, borrowed from chapter 15, where it lives in full. Infrastructure has sorted the sexes before. The plough needed upper-body strength and could not be worked with an infant on the hip, and the descendants of plough cultures still have fewer women in work, in business and in politics today, a pattern that follows the children of migrants into countries that never used one. The tool is gone. The sorting outlived it by centuries. A city is a tool too, and at Aspern a town's worth of people are about to move into one. Whichever passenger the route is drawn for is the passenger the route keeps producing.

So the three questions, at this scale. What is measurable: the queue, the pram on the stair, the extra minute at the cubicle door. What is statistical: the loop, its quarter of men's trips and its third of women's, and the lonely man on a low income whom nobody's plan was drawn for. What is ours: whom the line is drawn for. The route is a line somebody drew, and a line somebody drew can be drawn again. Here is the pencil.

THE STRONGEST CASE AGAINST

The chapter argued that cities were drawn for one passenger and that changing the fence, the fare and the lamp lowers the friction on the carer. Here is the best case that it has argued from one city's self-portrait and a handful of surveys.

Start with the lamp, which every city reaches for first. Paul Marchant spent the best part of a decade showing that the fifth in the lighting review is what regression to the mean looks like when councils light the streets that just had a bad year, and then the English and Welsh lighting cuts were checked against crime and nothing was found. The chapter's reply was to retreat from crime to fear. That is honest, and it is also a retreat. Fear is measured by asking, and a city that lights a street and then asks its women whether they feel safer is marking its own homework.

Then Vienna. The entire city-scale case rests on Vienna's account of Vienna: the survey at second hand, the park outcome on the city's word, the Aspern poll a poll of people who chose to live there, published by the developer. No one has compared the women of Vienna with the women of a city that did nothing and found that they travel faster, work more or fear less. Decades of gender mainstreaming may have done a great deal. On the evidence in front of me, it has also not been tested.

Third, the loop is not a sex difference at all. Janet Hyde's great synthesis found most psychological sex differences small or near zero, and the travel surveys are no exception: a quarter of men's trips in Montréal are care trips, and the father with the scooter makes the mother's loop. If the loop is a carer's loop, then the chapter is about care, not about sex, and a book about the Endocrine Accent has wandered into someone else's subject. And a city that builds the loop into its fares and courtyards for the people who walk it now makes the loop comfortable, and comfortable arrangements persist. The plough is the proof.

Last, the lonely city. The people the chapter found most alone are men on low incomes, and no fence, fare or lamp is addressed to them. A chapter that opens with a woman on a bus and ends with a man on a low income has changed the subject, and should say which one it is about.

IN PRACTICE

If you sit on a council, draw a timetable, run a housing office or only ride the bus, this is what the evidence supports at the city's scale, in conditionals, with the places it fails.

If your travel survey files the nursery run under other, then you do not know who rides your network and you cannot plan for them, so count the loop first, as Montréal and Washington did, before you change a single route. The failure mode is that counting is not building. A figure in a survey tends to stay in the survey.

If your fares charge by the boarding, then a morning of short hops costs several times a single commute and the person paying several times is the carer, so charge by time and let the transfer cover the loop. The failure mode is revenue and fraud control, and the evidence is an inference from fare-card data, not a trial of the alternative.

If your network is a star, then run the orbital route the survey shows and accept that it loads lightly. The failure mode is the empty bus, which a transport authority tends to read as a failed route rather than a served one.

If your building code gives the sexes equal floor space for toilets, then the women's queue is a mathematical certainty, so either set the fixture ratio in women's favour or take the sign off a room of enclosed cubicles. The failure mode is acceptance: the simulation is a model, whether people in your city use a shared room at all is an open question, and men's waits can lengthen slightly if the urinals go.

If your parks have a fenced pitch with one way in, then the pitch tends to belong to whoever holds it, and the evidence from Vienna is that the ones who leave are the older girls. So open the fence on several sides and put the things they asked for along the edges, as Einsiedlerpark did. The failure mode is that the boys lose a pitch and say so, and that the evidence is Vienna's own case study with no outside evaluation behind it.

If you light a street, then say which problem you are lighting it for. For fear, the evidence is consistent and the fear is a real restriction on who goes out. For crime, the evidence is one review's reduction against another study's nothing, and a council that promises a crime reduction tends to be caught by its own statistics the following year.

If you subsidise housing, then make the subsidy conditional, as Vienna did at Aspern: pram storage, a lit stair without corners, a courtyard the kitchen looks onto, childcare on the ground floor. The failure mode is

Kern's. In a market without social housing the same features price out the carers they were built for, who commute in from the edge to serve them, which is the old line with a new name. And a satisfaction poll is a press release, not a comparison.

If you worry about loneliness, then notice who is lonely in your data before you plan for them. In Germany it is men on low incomes and people with a migration background, young and middle-aged more than old, and the planning literature on cities for women has nothing addressed to them. The thing a city can do that an app cannot is put two people in the same place by accident, which is the one thing the prediction model could not predict at all. So build the places where accidents happen, the bench, the park edge, the courtyard, and expect nothing in particular from any one of them. The failure mode is that nobody has run that trial, and that the lonely may not come.

And if the loop in your city is walked mostly by women, then build for the loop and not for the women, so that whoever walks it tomorrow finds the route already there. The failure mode is the plough: a city that makes one arrangement comfortable tends to keep it, and the person across the kitchen table may have wanted a different one.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 13

The Wall

It is a quarter to one on a weekday in a town in the west of Germany, and a woman is standing at a school gate with her car keys in one hand and nothing in the other. In a few minutes the doors open and the day's schooling is over. Not the lesson. The day. She has arranged her life around it the way you arrange a kitchen around the sink, which is to say without ever quite deciding to. Behind her are other women with other keys, and one man, who is, in the way of these things, a little conspicuous.

That evening she sits at the kitchen table with a tax form. The form asks for her husband's income and for hers, adds them together, divides the sum by two, looks up the tax on one half, and doubles it. Four operations. They were written into West German law in 1958, the year my mother turned twelve, and they are still there, which means they outlived her.

On the other side of a border that no longer exists, inside the same language and the same Sunday cake, a woman of the same age grew up in a country that took the opposite bet on all of this, for forty years. The crèche was down the road and open all day. The tax office did not care who in the household earned the money. Nobody stood at a gate at a quarter to one, because there was no gate. And the word for a woman who stayed at home was not raven mother. It was parasite.

Then in 1990 the border came down and the western rules were laid over the whole country: the same tax form, the same timetable, the same word. If behaviour followed the rules, the two women should have become one within a few years. The records say they did not. A generation

on, the eastern mother is still at work about 350 hours a year more than the western one. Call it a month and a half of working days.

There was a wall. It was made of concrete, and it was also made of a timetable and a tax table, and only one of the three came down.

What the West built

The Preface told you about the cover. Let me tell you what was on the other side of it.

Until the late 1970s the legal name for a West German marriage was the *Hausfrauenehe*, the housewife marriage. A wife could take a job only with her husband's consent, and the consent could be withdrawn if the job got in the way of what the law called her duties in the home. Abortion was a crime under Paragraph 218 of the criminal code, with prison as the sentence. These were the two walls the movement my mother's generation joined could see, and it went at the second one first, because the body was where the argument could not be lost.

Alice Schwarzer, a journalist who had seen the French manifesto that spring, gathered the women, Romy Schneider among them, and put their faces and their confession on the cover of *Stern* in June 1971. In 1974 the government passed a law making abortion legal in the first three months. In 1975 the Federal Constitutional Court struck it down, reading into the Basic Law a right to life for the unborn that the state was bound to protect, against the mother if need be. The German Democratic Republic, three years earlier, had legalised abortion on demand within a similar limit. The compromise of 1993, after reunification, is the one still in force: abortion remains an offence under Paragraph 218 and goes unpunished within that limit if the woman has been counselled and has waited.

Count what was won. The right to the body, in the compromised form a court allowed. The partnership marriage, in 1977, when the housewife marriage was struck from the law and a wife no longer needed anyone's permission to work. Both were fought for against the memory of a state that had treated women as a means of producing children, and that memory explains the movement's one great blind spot.

West German feminists did not, on the whole, demand public childcare. A crèche run by the state looked, from 1971, like something the pre-

vious regime would have built, or the one across the border was building, a place where children were collectivised and mothers made interchangeable. So the movement fought for the right to work and not for the means to. It won the right. And a woman who used it found herself alone against a tax code, a timetable and a word, with no vanguard behind her on the one question that decided her hours. Emancipation, in the West, became the double burden: a part-time job in the morning and the whole of the house for the rest of the day, because the men of the house had not been part of the argument and did not change.

None of it was German in origin. The idea that a home is a sphere apart from work, run by a woman whose idleness proves her husband's standing, was written in English, in Birmingham and Essex, between 1780 and 1850, in sermons, wills and house plans. Germany had not invented the breadwinner. It had written him into the tax code more carefully than anyone else.

Back to the four operations, because the arithmetic is the mechanism. Take a couple in which he earns a great deal and she earns nothing. Pool the incomes, halve the pool, and the tax schedule now sees two moderate earners instead of one large one. Because the schedule is progressive, two moderate tax bills come to less than one large one, and the saving is largest exactly when the gap between the two incomes is largest, which is to say when one of them is zero. Now let her take a job. Every euro she earns is added to a pool that is already sitting at his top rate, so her first euro is taxed as if it were his last. The couple's advantage shrinks with every hour she works. A microsimulation in 2004 found this one rule to be a primary driver of the exceptionally low full-time employment of married women in Germany.

Nobody is forced to do anything by this. That is the point. A tax code of this kind does not forbid. It prices. It prices her hour of work below his, in a schedule that neither of them wrote and few of them can explain. Add the timetable: schools that ended at lunch, and, for the under-threes, public care that was for practical purposes a rumour. The state's own infrastructure assumed somebody would be at home in the afternoon, and so it required it. The two rules did not need each other. They fitted, the way a sink fits a kitchen.

Then the word. A *Rabenmutter* is a raven mother, cold, a bird that leaves its young in the nest, and the word did what the tax code and the

timetable could not: it turned a logistical problem into a defect of character. A woman who put a small child in care was not short of options. She was short of love.

A word about layers, because this is where the book's rule earns its keep. The tax code is culture: it was voted on, and it could be voted away. Its effect is statistical: not every married woman in the West stayed home, and the ones who did not were paying a price, not breaking a law. And the thing being taxed was a body, the one in the household that could gestate and lactate, which the state, with the best Catholic intentions and the worst arithmetic, decided to shield from the market by making the market expensive for it.

The compounding is the part my parents' generation did not see. Older women in Germany draw about half the pension income of older men, because the pension follows lifetime earnings and lifetime earnings follow the hours dropped at the first birth. The welfare state that was built to protect the family from the market protects it, in the end, by making the mother dependent on it in old age. Nobody designed that either.

What the East built

The East did not start from a theory of the family. It started from a shortage of hands.

The German Democratic Republic lost workers to the war and then to the border, which was open until 1961, and a planned economy that cannot find labour has an interest in women that a market economy can afford to lose. The constitution of 1949 wrote in the right to work and to equal pay. Public childcare was built, and by the late 1980s it covered nearly every child in the country. By 1989, the last full year before the wall came down, nine women of working age in ten were in the labour force in the East, against a little over half in the West.

Then the planners noticed something a planner would notice. The women were working and they were not having children. Fertility fell, and from the 1970s the state turned to what Germans on both sides later called *Muttipolitik*, mummy politics: policy designed not to free women from motherhood but to make it compatible with the plan. The flagship was the *Babyjahr* of 1976, a paid year off after a birth, the job held open, with

the option of reduced hours afterwards and no penalty on the way back. Note the name. The year went to her. The reduced hours went to her. The state that had abolished the breadwinner's tax advantage had not abolished the idea that the baby was hers, and that is going to matter in a few pages.

And the word ran the other way. A woman who stayed home with her children was a *Schmarotzer*, a parasite on the collective, somebody avoiding her civic duty. Two states, one language, and each had found a single noun to put on the woman who did what the other state required.

Inside the flat, nothing like this symmetry held. The East put women into the shift and the office and left the kitchen to them as well. Its own surveys show full-time paid work plus near-total responsibility for the home, the double burden, and men whose domestic hours did not move to meet their wives' paid ones. The socialist state did not challenge male privilege at home. The emancipation of East German women was a labour policy, delivered from above, and a labour policy does not need to redecorate the kitchen.

So when the two movements met, in the early 1990s, each saw a stranger. Western feminists looked at the East's Independent Women's Association and saw obedient subjects of an authoritarian state. Eastern women saw people who had argued about the body for twenty years while they ran a household on a full-time wage, and who now stood by while the crèches and the *Babyjahr* were dismantled. Both were right about the other. Neither had the other half.

What the records say

It is now thirty-five years since the rules were made the same, and the best part of the experiment is that German bureaucracy kept the receipts.

Barbara Boelmann, Anna Raute and Uta Schönberg, three economists, took an enormous random sample of the country's social-security records and asked them a careful question. Not simply how much mothers work on each side of the old border, but what a couple is paying for when a mother works: her wage, the childcare bought, and the social cost of being seen to work, all in one model. The headline you know from the

Preface. The eastern mother works more. The fathers, East and West, the same.

Wages are lower in the East and childcare is dearer in the West, so a naive reading says the East works more because it is paid less and charged less. The model says the opposite. Equalise the wages across the country and the gap does not close. It nearly doubles. The higher western wage is, in effect, what the market pays a western mother for working against the word, and take it away and more of her stays home. The wage was holding the western figure *up*. Their estimate, and it is the hinge of the chapter, is that the working-mother stigma alone accounts for 73 per cent of the gap.

Then the movers. West German mothers who moved east came, within a few years, to work like East German mothers, and the ones who later moved back west kept doing it. The behaviour did not depend on the crèche being outside the door. It had been learned and it travelled. And in western firms where the share of eastern colleagues rose, western mothers went back to work after a birth sooner. You can catch the East at the office.

Two German states ran the same people through two arrangements for forty years, and the difference in what mothers do is still measurable a generation later.

The economists who measure what a first child does to a woman's earnings, the child penalty that chapter 8 put at about 60 per cent for a German mother, turned the same instrument on the old border in 2024 and found that mothers raised in the East carry a smaller penalty than mothers raised in the West, whichever side of the line they now live on. It is where you grew up, not where you live.

Quentin Lippmann, Alexandre Georgieff and Claudia Senik went inside the household. Their data is the country's long-running household panel, and their question is what a couple does in the year after the wife's income crosses half the household total. The Preface told you the answer: in the West she does more housework, is more likely to leave her job, and the couple is more likely to divorce, and in the East none of it. In 1992 nearly three in ten East German wives out-earned their husbands, close to twice the western share, and the gap has not closed. The reaction sociologists call doing gender, the restoring of a hierarchy the payslip has just upset, is not a human universal. It is a western one, in the German sense of western, and forty years of a different tax office was enough to switch

it off.

Two more records. Put attitude questions to both halves of the country and East Germans come out markedly more likely to hold egalitarian views on sex roles. And the register of births. In the West of my mother's childhood a child born outside marriage was a rare exception, and getting rarer. In many eastern districts today it is the ordinary case. Demographers who traced the divide back found the East ahead already before the division, and the GDR then made the lead permanent.

Put the records together and they say one thing clearly. Give a population a tax code, a timetable and a word for forty years, and you move its mothers' hours, its wives' willingness to out-earn their husbands, its attitudes and its birth register, and the movement outlives the regime by a generation and travels in a suitcase. That is the culture layer, measured. It is the largest effect in this book.

What the state builds, the crèche, the school day and the tax table, decides more of a mother's working life than anything in her body.

Where the wall stops

Now the other column.

The fathers did not move. In the records that found the gap between the mothers, there is no gap between the fathers at all. The East had abolished the breadwinner norm, pushed nearly every woman of working age into paid work, and left the men's hours exactly where the West had left them. What reunification produced in both halves, the sociologists who tallied it concluded, was a convergence not on the dual-earner household but on a modernised breadwinner: he full-time, she part-time. The East had been told, for forty years, what a mother was for. Nobody had told it what a father was for, and so a father went on being what he had been in 1949.

The kitchen did not move. A state can move public behaviour, hours in the shift and the office, and meets something close to insurmountable friction at the private negotiation between a man and a woman in a flat. The East's double burden was the proof, and reunified Germany's family panel repeats it in both halves: the laundry and the cooking still mostly hers, and the symmetrical couple reverting to the old split at the first birth

almost invariably. Chapter 7 has those figures. Whether the eastern version of equality raised women's satisfaction or merely moved their tiredness from the home to the factory is an open question, and I do not know.

The young did not move, or stopped moving. The best cohort study took the German General Social Survey from 1982 to 2016 and separated the effect of age from that of the year from that of the birth cohort. In the West the drift towards egalitarian attitudes stopped for men born around 1956 and later and for women born after 1966. In the East the younger cohorts also soften a little, from a higher level. I was born in 1975. The men of my cohort, in the West, are the first in forty years who are no more egalitarian than their fathers, and I say that about my own side of the argument before anyone else can.

And here are the two readings of the stop, and the chapter has to carry both. The first is Catherine Hakim's, from 2000: that once a society gives women contraception, equal law and white-collar work, they sort themselves by preference, a minority who put the home first, a minority who put work first, and a majority in between who adapt and prefer part-time. On that reading the West's part-time mother is not a victim of a tax code. She is the majority, and the East was a forty-year interruption to her. The second reading comes from a survey experiment in 2020 with the title "Money Matters!": show people a scene in which a mother's wage is what keeps the household afloat and the belief that mothers must be at home collapses on the spot. Preferences, on that reading, are what the price list looks like from inside.

I think the records settle the first half of this and not the second. A preference that moves when you cross a border, keeps moving when you cross back, and spreads to the next desk at the office is not a preference in Hakim's sense. It is a price. But a father whose hours do not move in either direction, under either regime, is not a price either, or not one anybody has yet tried to change, and the years since Germany introduced an earnings-related parental benefit with months reserved for fathers, the *Elterngeld* of 2007, are too few for the records to say whether that was the missing crane.

So here is the tool, and it is a reading instrument rather than a lever. The wall shows you how to separate three things that every argument about men and women runs together. What moved under the regime is culture, and it is large, and it is the mothers. What did not move under

either regime is either the body or a wall nobody has yet tried to take down, and the fathers and the kitchen are on that side of the line, and the records cannot yet tell which. And what the regime was never asked about, the accent in the room when the hours are finally equal, is the question the next three chapters and the last one take up.

The concrete wall came down in a year. The timetable and the tax table did not. The wall through the kitchen, nobody has yet tried.

THE STRONGEST CASE AGAINST

The case against comes from both ends of the bookshop.

From the evolutionary shelf, the kibbutz. Lionel Tiger and Joseph Shepher went through three generations of occupational records on the Israeli kibbutzim, where childcare, laundry and cooking were collectivised, and found that by the second and third generations about three quarters of the women had gravitated to education, health and services and about three quarters of the men to agriculture, industry and management. The critics say the sample began small and the kibbutz sat inside a patriarchal region. But the shape is the opposite of the East German shape. Remove the price and the division comes back. Hakim's majority, on that reading, were the kibbutz-born generation, and the East's working mothers were a wage bill.

From the same shelf, the paradox chapter 5 carries: measure economic preferences across dozens of countries and the gap between the sexes is largest in the richest and most equal places, a finding fought over since. If it holds even in part, the German drift towards a modernised breadwinner after 1990 is not only the West's tax code winning. It is what freedom looks like, and the East's dual-earner household was the thing that needed a state to hold it up.

From the constructivist shelf, the harder blow, because it is aimed at the half of the chapter I like. The fathers did not move, I said, so perhaps the body is in the way. The opponent's answer is that nobody pushed them. A still father in a regime designed around mothers is not evidence of an accent. It is evidence of a regime designed around mothers. And the clinching comparison is Japan, where a woman does more than five times the unpaid work of a man, the widest gap in the OECD, in the country

where men's paid day is the longest. Where men are not in the house, there is no private negotiation to lose. The kitchen does not stay with the woman because of her body. It stays with her because he is at the office, and the office is a tax code too.

I cannot refute either shelf from the German records alone. The kibbutz says the division returns when the price is removed. The East says it stays removed when the regime is gone. They disagree. That is not a reason to pick one. It is the subject.

IN PRACTICE

The scale here is the state, and most of you do not run one, so first what a person does inside one, then what a state does if it is listening.

If your country splits a couple's income for tax, then the arithmetic is already in your kitchen whether or not you have read it, and the first thing to do is read it: work out what the second earner's first hour is worth after the schedule has taken his rate off it, and decide with that number on the table rather than with the gross. The failure mode is reading the number and treating it as nature. It is a line in a statute, it was passed in 1958 against other options, and Scandinavia chose the other options.

If the school day ends at lunch and the crèche is a rumour, then one of you is going to be at the gate at a quarter to one, and in the German records it is her, and the records say the reason is the price list, not the person. Decide who stands there before the first birth and not after it, because the family panel says the symmetrical couple reverts at the birth almost invariably. The failure mode is the plan that assumes the timetable is negotiable. It is not. Only the person at the gate is.

If you run a firm in a breadwinner country, then your office is a border. Western mothers in firms where the share of eastern colleagues rose went back to work sooner after a birth, which means the composition of a team is a family policy whether or not you meant it as one. The failure mode is treating this as a lever: the records give the direction and not the size, and an engineered peer effect is a different thing from one that happens.

Now the state.

If you want mothers in paid work, then the records say three things move them: the price of her hour, the length of the school day and the

word for her, and the word alone accounts for most of the gap. A tax reform without the crèche tends to move little, a crèche without the word tends to move something, and the East moved everything, with methods no democracy should want. The failure mode is believing that minds change on their own once the law is equal. The cohort data say they stall.

If you want fathers in the house, then no German regime has yet shown you how, and the honest instruction is that nobody has tried with the force that was used on mothers. The *Elterngeld* with its reserved months dates from 2007 and the records cannot yet say what it has done. The counter-evidence is the Japanese column: where men's paid hours are the longest, no family policy reaches the kitchen because the man is not in it. A father's hours at home are set at the office, and a state that wants to move them has to move the office first.

If you want the private sphere to change at all, then the records say the state's reach stops at the front door in both regimes, and the kibbutz says it stops there even when the front door has been removed. What moved the East's kitchens, to the extent they moved, was not a decree. It was that she out-earned him and nobody reached for the ironing board. Give couples that fact, through the tax code, and the rest is theirs, which is the only place it was ever going to be decided.

And if you are tempted to run the experiment again, in either direction, then remember what it cost. The East's equality was a labour policy with a wall around it. The West's family was a tax advantage with a word attached. Neither was chosen by the women it was built for, who would, given for once the whole question, have chosen differently from both. The wall through the kitchen is a design, and designs can be changed, slowly, by the people who notice they are standing in one.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 14

Queens at War

A woman is standing up in a village council meeting in West Bengal, and the room is not used to it. In 1993 India changed its constitution so that one chair in three at the head of a village council would be reserved for a woman, and which villages got one was decided at random, so that nobody could later say the places with women in charge were the places that had wanted them. She is reading out the budget. More of it than last year goes on drinking water, which is what the village's women complain about most when anyone asks. Along the wall sit those women, who in an ordinary year get through the whole meeting without a word, and this year they are speaking up about twice as often as they do when a man has the chair. Nobody told them to. The chair changed, and the room changed with it.

Hold that picture, and a second one beside it. Somewhere in early modern Europe a queen is at war. Not defending. Attacking. Her husband is running the treasury and the army while she runs the state, and together they are a more dangerous neighbour than any king whose consort is kept out of the room. Across every reign that two economists could find in Europe between 1480 and 1913, a state was markedly more likely to be at war in a given year under a queen than under a king.

Both pictures come from the same kind of evidence: the lottery, the accident of birth, the thing that decides who holds power without asking anyone's opinion of women. It is the only kind this chapter trusts, because the countries which put women in charge are already different from the countries which do not. And the queen is the chapter's standing image

because she breaks the folk theory on both shelves at once. She is not the gentler sex in a crown. She is not a man in a dress either. She is a person with a different set of constraints, and the constraints, it turns out, are most of the story.

The lottery

The village councils matter because of the lottery. Raghabendra Chattopadhyay and Esther Duflo, economists, took village councils in West Bengal and Rajasthan after the reservation came in and compared the ones that had drawn a woman chair with the ones that had not. Where women's complaints ran to drinking water, the women's councils spent more on water. And the women of the village spoke at the meetings about twice as often. That is the finding I would keep if I could keep only one, because it is about the room, not the chair.

What happens when the lottery ends? In Mumbai, in wards where a reserved term had lapsed, a woman was about five times as likely to win the next open election as in a ward that had never been reserved. The quota did not just put a woman in office for a term. It let the voters find out what a woman in office looks like, and the parties that she can win.

Now scale up. Take the Indian state legislatures and ask what a larger share of women in the assembly does to the children of the state. The answer is that newborn deaths fall by about a third, through more antenatal visits, more iron for the mothers and more births in a clinic. Take it global, across the developing countries that adopted quotas for women in parliament, and after a quota maternal deaths fall by about a tenth and girls stay in school longer.

So the first finding is as solid as anything in this book. Change who holds the chair and the budget moves towards water, clinics and children. A parliament with more women in it does not become gentle. It becomes different in what it spends on.

One more from the Indian file, because it answers the obvious objection. If women spend on health and schools, do they neglect the economy? Take the close elections to Indian state assemblies, the ones a woman narrowly won against a man or narrowly lost, so that the winner's sex was as good as a coin toss. Constituencies that got a woman grew faster,

measured by the light visible from satellites at night, the economist's way of counting without asking. The men in the same close races were far more likely to be carrying criminal charges, and while in office they grew rich faster than the women did.

That last line sounds like a verdict. It is not one, and the next section is about why.

The fairer sex, audited

Here is the finding that launched a thousand columns. Put the share of women in a country's parliament next to its perceived corruption and the two move together: more women, less corruption. The paper that found it, in 2001, asked in its title whether women were the fairer sex, and a great many people answered yes before reading it.

The answer that survived is more interesting. The relationship holds in democracies and vanishes in autocracies, which it should not do if women were cleaner by nature. Look closer and women reduce corruption only where the odds of getting caught and punished at the ballot box are high. Two things do the work. One is statistical: women on average take fewer risks, and corruption in an accountable democracy is a risk. The other is cultural, and it is the sharper of the two. Voters hold women to a higher ethical standard, so a woman caught in a scandal is punished harder at the polls than a man caught in the same one. Women in politics are cleaner, where they are cleaner, partly because they are watched more closely. That is a fact about the audience, not the molecule.

And the risk aversion that does half the explaining is a smaller thing than the folklore. The best audit of the laboratory games that find women more cautious found that the studies showing a difference had been printed and cited and the ones showing none quietly filed, and once you allow for which papers made it into print the gap shrinks a good deal. The variation among women, and among men, dwarfs the variation between them. The idea that women's social preferences bend more to context went to a large replication and failed it. Each of these is a finding about averages, with an overlap you could drive a bus through.

One more filter stands between the population and the podium, and the folk theory always forgets it. The women who reach the top of na-

tional politics have come through the same selection machine as the men, and the ones who get there sit in the tails of every distribution, hard to tell from their male colleagues on competitiveness, risk tolerance and the appetite for dominance. The central bankers in chapter 10 were the clearest case, with the women on the committees voting for *higher* rates than the men. An average is a fact about a population. The woman at the dispatch box is one person, and she got there by not being average.

Which brings us to the queens.

Queens

Oeindrila Dube and S. P. Harish, economists, wanted a case where women held supreme power and nobody had voted for them, so that the countries which choose women could not be confused with the women they choose. Hereditary monarchy is that case. They assembled every reign they could find across the kingdoms and principalities of Europe between 1480 and 1913 and asked a plain question: is a state more or less likely to be at war in a given year when a queen holds the throne? Even here selection could creep in, since a queen tends to inherit when there is no brother. So they leaned on two accidents that nobody chooses, whether the previous monarch had a first-born son and whether the previous monarch had a sister, to isolate the part of a queen's accession that is pure luck of birth.

The answer was 39 percentage points. A state under a queen was that much more likely to be at war in any given year than a state under a king. Then the finding splits in two, and the split is where the chapter earns its keep. Unmarried queens were more likely to be *attacked*: their neighbours seem to have read a woman alone on a throne as an opening. Married queens were more likely to be the aggressors, and more so than married kings. Dube and Harish's explanation is a division of labour. A married queen could put her husband in charge of the army and the exchequer and keep the state for herself, in effect running a dual executive, which gave her both more capacity for war and more allies to wage it with. A married king, in the same centuries, did not hand his wife the treasury. The queen was not more warlike as a woman. She was more thoroughly staffed.

Queens went to war more often than kings, and the reason is in the arrangement around the throne, not in the woman.

That is the most useful single result in this chapter, because of what it does to both shelves. To the shelf that says women in power would end war, it says: here are four centuries of women in power, and they fought more. To the shelf that says women cannot or do not wage war, it says the same thing. And to this book it says that the constraint is the variable. A skeleton is measurable. An appetite for risk is statistical. The arrangement around a throne is neither, it is ours, and it is the one that decided the wars. Give a leader a capable co-executive and the state fights. Take the constraint away and the sex of the person in the chair tells you very little about whether the cannon is loaded.

Set that beside the finding that has become the slogan of a movement. Valerie Hudson and her colleagues built a database of women's lives across most of the world's states, the WomanStats project, and found that the physical security of a state's women predicts that state's peacefulness better than its wealth, its democracy or its religion. The best single predictor of a state's violence abroad is how it treats its women at home. The objection is the one you would expect: rich, stable, democratic states both protect their women and avoid wars, so the arrow could point either way, or both ways from a third thing. The finding has been fought over since, and it has not fallen.

I take the objection seriously and carry the finding anyway, because of what Dube and Harish add to it. If women in power are not the peaceful sex, then the Hudson pattern cannot be about the temperament of women. It has to be about what a state is like when it treats half its people as property. That is a claim about institutions, and institutions can be changed.

What kind of state is that? The deep-history databases that chapter 2 used for the ledger of who cooks have an answer. Across the standard sample of the world's societies that anthropologists have worked from since 1969, there is not one in which women hold the military leadership. Peggy Reeves Sanday scaled a large part of that sample in 1981 by women's economic control, political voice and symbolic standing, and found a pattern she called inner and outer orientation. Societies that lived off gathered plants and had long stretches of calm tended to share power. Societies under stress, in famine, on the move or at endemic war, came

to rely on the two things where the larger body counts most, hunting large animals and fighting, and the men who did those things took the rest. In Sanday's data that outer orientation travels with high rates of violence and with a culture in which rape is common, a pattern others had already called the male supremacy complex: a society that needs young men to die for it pays them in extra wives, a monopoly on politics and the subordination of its women. And the arrow can run the other way. Although most societies in the ethnographic record permitted a man more than one wife, the rule of one spouse each spread because of what it did to the group: fewer unmarried men without prospects, less murder, rape and robbery, and fathers who invested in children rather than in the next wife.

Which is the point at which this chapter should turn to sex ratios, surplus men and the stability of states, and cannot.

The section on sex ratios, polygyny and social stability is not yet written. Its research report is missing from the library.

Whatever the gap turns out to hold, the rest of the evidence does not run through female gentleness. The best test of this took a sample of preindustrial societies in 1978 and checked the assumption underneath every "status of women" index ever built, that a society which gives women one kind of standing gives them the others. It does not. Women's standing comes in separate pieces, property, domestic authority, ritual inclusion, sexual freedom, political voice and the rest, and the pieces barely move together. A society could give women the home and the temple and keep them out of the council and the war band entirely. There is no single thing called the position of women that rises and falls as one. There are many, and a state can be generous on most of them and murderous on the rest.

Two columns in the statute book

Now turn the question around. Not what women do to the state, but what the state still does to the sexes, after decades of laws that promise to treat them alike. The law has two columns, and some of them are written in bone.

Start with the one that cannot be legislated away. Norway, in 2013, passed a law conscripting women as well as men, the first member of

NATO to do so. The armed forces put men and women in the same dormitories, and a researcher who went to the camps found that shared rooms produced a sibling-like team feeling and less harassment, while the all-female rooms had more drama. A cheerful result. Then the body, which does not read Norwegian statutes. In Israel, where the Anti-Aircraft Forces put men and women through identical basic training, the researchers counted the stress fractures: about twice the rate in the women as in the men, and the women's fractures sat in the shin, the thigh and the pelvis, the bones an infantry soldier walks on. In light infantry the same team later found nine times the rate. Lower bone density and smaller muscle mass are measurable. The law can make the duty identical. It cannot make the skeleton identical, and a state that pretends otherwise breaks its own soldiers. And still, as the same Israeli data show, individual women with high baseline fitness complete the tours, which is the overlap doing its usual work against the average.

The second column is written in culture, and it runs the other way. The best study of sentencing took the American federal criminal cases and followed each one from arrest to sentence, so that the leniency at every stage could be added up. Holding the arrest offence and the criminal history constant, men received sentences 63 per cent longer than women, and women were far likelier to avoid prison altogether. The whole gap was several times the size of the racial gap found the same way. France, where every conviction over four years sat in one dataset, gives the finding its cleanest form. Among pairs of co-offenders, one man and one woman, convicted of the same crime by the same judge on the same day, the man got the longer sentence many times more often than the woman did. More women on the bench shrank the gap. The paternalism is mostly the men's, and it is an average, women commit far less violent crime, applied to one defendant, which is exactly what this book keeps saying not to do.

The family court keeps the same two columns. The law's default has been the mother's house in the large majority of contested cases, and the evidence has been pushing against that default for a generation: children in joint custody come out closer to children of intact families, their parents describe less conflict, not more, and when courts began denying overnight stays to fathers of infants a large consensus of researchers said the evidence did not support it. Chapter 8 had the physiology of the father's brain. Here is the law catching up with it.

So when does a mixed body decide differently from a single-sex one? Chapter 10 found the laboratory answer unstable. The state-scale answer is steadier, and it comes with a mechanism from the animals. Two researchers who model how animal groups decide showed that when the members have different interests, despotic decisions, one animal deciding for all, cost the group more than shared ones, because a shared decision lands nearer the middle of the group's interests and avoids the extremes. That is the whole case for a mixed body in one sentence: not that women are wiser but that two distributions averaged together have a shorter tail than one. Where the stakes are highest the human evidence runs the same way: more women in a parliament goes with lower carbon emissions, and in the first wave of the pandemic of 2020 the countries led by women locked down earlier and lost fewer people. And in the American death-penalty cases where the jurors were interviewed afterwards, white male jurors were the readiest to sentence a defendant to death and the women on the jury weighed the mitigating evidence more heavily, so that a mixed jury kills less often than a male one.

Here is the turn. In the nineteenth century men held every vote and every seat, and they gave the vote and the seats away. Matthias Doepke and Michèle Tertilt asked why, and their answer is in the title of their paper: "Women's Rights: What's in it for Men?" As the return on education rose, a man's best bet for his own line was to have his children schooled, and women spend a larger share of household money on children than men do, so men voted to give women property rights, and in time the vote, to get the money where they wanted it. It is not a romantic account of emancipation. It is a true one, as far as it goes, and it hands you the tool for this chapter's scale.

Look at what the two columns have in common with the kitchen table. The state assumes one body when it writes the physical standard, and breaks the other. It sentences one person by the distribution of her sex. It hands the children to the parent the folklore expects to want them. The state is the couple's arrangement at scale, and it inherits the couple's design faults. Nobody at the top of a state has ever shared power because they were asked nicely. They share it when it pays, or when a lottery makes them, or when a queen with a capable husband is on the border. If you want a state to behave differently towards its women, do not argue about what women are like. Change what it costs to treat them badly.

THE STRONGEST CASE AGAINST

The chapter has argued that women in power change what states do, and that the sex of the person in the chair says little about war. The case against runs through both halves, and it is about confounding.

Begin with the corporate boards, the best-measured bodies on earth, where the effect is close to nothing. Pool the studies of women on boards and you get a small relation with accounting returns and none with market performance or with risk. The Norwegian quota boards laid off fewer workers and earned less in the short run, which is a priority, not a competence. If sex composition does so little where the data are best, the burden is on the political findings to show they are not the same thing seen through a thinner lens.

And the political findings have an identification problem that the Indian lotteries solve and nothing else does. Leah Windsor and her colleagues made the point about the pandemic and the climate: the countries that elect women are high-trust, low-hierarchy countries, and those are the countries that lock down early and tax carbon, whoever is in the chair. The Hudson pattern is open to the same reading, and her own critics say so.

Then the averages themselves may be thinner than the chapter let on. The publication-bias audit is not a footnote: it means the risk-aversion gap that half the corruption story rests on was partly manufactured by the journals. And the one bodily finding in women's favour, tend-and-befriend, rests on oxytocin, which the Preface has already recorded does not do what it was sold as doing. Strip those out and what remains is: women in office spend differently. That is enough for the chapter's first half. It is not evidence of a different nature.

Finally the queens could be read against me. Thirty-nine points is a vast effect from one continent's reigns, and the husband as co-executive is an inference about staffing from the pattern of who attacked whom. A sceptic could say the finding is about early modern Europe and tells us nothing about a cabinet. She would be right that it tells us nothing about women. It was never about them. It tells us about constraints, and if constraints are the story then the chapter's critics and I are closer than either of us would like.

IN PRACTICE

This chapter's scale is the state, which most readers do not run but all of them vote in, pay for, are sentenced or conscripted by, and sit on the juries and councils of. The conditionals are for that reader, and for the few who draft the rules.

If you are deciding whether a quota is worth the cost, then the Indian evidence says the gain is not mainly the term the quota buys but the term after, when a woman is five times as likely to win the open seat. The failure mode is a quota without an exit: the effect was measured where reservations lapsed, and a seat reserved forever teaches voters nothing. The counter-evidence is the token seat: a quota that puts a faction's women in the chair tends to buy the faction, not the change.

If the argument for women in office is that they are more honest, then stop making it, because it holds only where the press is free and the next election is close, and it tends to collapse the moment the accountability does. The failure mode is the one the voters supply: the honest-woman story raises the price of her first scandal, so the argument that put her in tends to be the argument that throws her out.

If you are reading a correlation between women's rights and peace, then read it as a fact about the state that violates its women, not as a promise about the women it would elect. The queens are the counter-evidence, four centuries of it. The failure mode is the peace movement that puts a woman in charge of the war ministry and is surprised.

If your state writes the same physical standard for both sexes in a trade that loads the skeleton, then expect the Israeli numbers, roughly double the fractures in mixed training and nine times in light infantry, and build the training to the bone rather than to the statute. The failure mode on one side is a standard lowered for women, which breaks the women it admits and the trust of the men beside them. The failure mode on the other is a standard pretended to be sex-blind, which breaks women in the first months and calls it attrition. The counter-evidence is the individual woman who passes the light-infantry tour, and she exists in the same Israeli data, which is why the standard should be a standard and not a sex.

If your courts are handing men sentences 63 per cent longer for the same file, then that is the average, women commit less violent crime, applied to one man in a dock, and the cure is the same as everywhere in this

book: sentence the person, not the distribution. The failure mode is the correction that over-shoots and discovers that equality for women in the dock means more women in prison, which few of the people asking for it had in mind. The counter-evidence is the French bench: more women judges narrowed the gap, which suggests the leniency is a paternalism that a mixed body tends to dilute.

If you sit on a body that decides something irreversible, a jury, a board, a committee that goes to war, then the case for mixing it is the animal one, not Hudson's: two distributions averaged have a shorter tail than one. The failure mode is the token, the single woman in a room of men, who in chapter 10's teams tends to speak less and change nothing. The counter-evidence is the central bank, where the women were the hawks, so a mixed body tends to be less extreme in ways you did not predict and may not want.

If you want power shared at the top of a state, then ask what it costs the people holding it not to share, because that, in Doepke and Tertilt's account, is the only reason it has ever happened. The failure mode is the argument from virtue. It has a four-hundred-year record, and the queens are in it.

And if you were hoping this chapter would tell you that a world run by women would be a gentler one, then I don't know, and the best evidence I have says that a world in which it is safe to be a woman tends to be a gentler one, which is a different sentence, and the only one I can sign.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 15

The Long Argument

A girl of fifteen sits at a desk in a school in Finland, in the spring of 2018, with a booklet in front of her that pupils her age in dozens of countries are filling in that spring. The mathematics is done. Now come the questions about herself. Whether she enjoys reading. Whether, when she does badly at something, it is because she lacks the talent for it. She ticks a box. The booklet goes into a pile, the pile into a crate, the crate into a database with a four-letter name, and from there her tick travels to Paris, to be laid beside the ticks of girls in Tunis and Seoul and Lima and argued over by people she is never going to meet.

She has just joined the oldest argument in this book.

It was being conducted in Greek in the fifth century before Christ, in a treatise on the diseases of women. It was conducted in Latin by an abbess on the Rhine, in classical Chinese by a Tang doctor who decided that women needed their own medicine, and in Hebrew by rabbis who found that two categories were not enough for the bodies in front of them. It was conducted in law codes written on bamboo and in dowry contracts in Cairo and in the stone of a female pharaoh's temple, where she is carved with a beard. It is conducted now in American journals, in standard deviations, and it has not been settled, because the people having it have never agreed what it is about.

Someone, long ago, filed it. The file has been added to ever since, and then, not long ago, the whole thing was labelled reactionary and pushed to the back. The Preface promised that this book would open the drawer. The chapters so far have taken out one folder at a time. This chapter pulls

the drawer all the way out and sets it on the table.

Pull up a chair. It is heavier than you expect, and the oldest folder is at the bottom.

The oldest folder

The first people to write the argument down were doctors, and the first thing they wrote down was the obvious. The Hippocratic writers held that a woman's flesh was looser and spongier than a man's, that it took up more fluid, and that the monthly bleeding was the body's way of clearing the surplus. That is a claim about the body, made from looking, and the half of it that says women bleed and men do not has held up rather well.

The explanation did not. Aristotle decided that only the male contributed seed and the female contributed matter, as a carpenter gives shape to wood or rennet turns milk to curd, and that a female was what happened when the womb ran short of heat before the job was finished. An unfinished male. Galen, in the second century of our era, gave the woman a seed of her own, colder and weaker, and described her organs as the man's turned inward. For centuries after, the medicine of Europe, Byzantium and the Islamic world argued inside the space those two men had marked out.

In 1990 the historian Thomas Laqueur, in a book called *Making Sex*, told the story a different way. Until the eighteenth century, he said, Western medicine knew only one sex, the male, with the female a lesser version on the same ladder, and the modern idea of two distinct and opposite sexes was invented rather than discovered, when Enlightenment talk of equality left the old hierarchy needing a new foundation. The book became a founding text for the view that sex itself is a cultural artefact. The historians of medicine have spent the years since taking it apart. Anatomists had a distinctly female skeleton on paper, pelvis and all, by 1600, Felix Platter in 1583 and Caspar Bauhin a few years later, two centuries before the politics that Laqueur said required it. The subtler point matters more. Laqueur had treated menstruation as a minor local case of general bodily flux, when to every physician before the modern age it was the defining fact of what a woman was. The settled view among his critics is that

the metaphors changed and the observation did not. A doctor can call the womb an inverted penis and still know perfectly well which patients bleed.

That is the first thing the drawer teaches, and it is worth saying plainly. The old doctors saw most of what we see. What they got wrong was the filing.

So what did they see that holds? Three things, in three traditions. In China, in the seventh century, the physician Sun Simiao set women's medicine apart as a discipline of its own, and his reason was not cosmic but bodily: menstruation, pregnancy, birth and recovery were strains the male body never met and needed their own pharmacy. The decoction he prescribed for painful periods is still prescribed. In Salerno, the collection known as the *Trotula* turned out, when a modern editor took it apart, to be three texts, one of them the work of a real woman, Trota, whose treatments are practical, almost untouched by the theory of humours, and who held that infertility could lie in the man's seed as often as in the woman's womb. And on the Rhine in the twelfth century Hildegard of Bingen, abbess and polymath, wrote of a woman's pleasure as part of conception and then did something nobody had done before her. She described a phlegmatic man and a sanguine woman, temperaments that crossed the sex line, and a varying vital sap she called *viriditas* that made each body a point on a range rather than a type. An abbess, in the vocabulary available to her, had drawn the Overlapping Bells. Ibn Sina, in the eleventh century, held that there were two seeds, which is closer to the truth than the ladder.

Two of the old observations can be checked against modern numbers, and both pass, with a caveat each. The old texts noticed that women's moods and bodies shifted before bleeding. A modern review that pooled studies from around the world finds premenstrual syndrome everywhere, and finds it at 12 per cent in France and 98 per cent in Iran. A condition of the body is being reported through a cultural filter so coarse that the filter is a finding in its own right. The old texts also noticed that a nursing mother did not conceive. The best modern review finds that she mostly does not, for the first six months, if the baby takes nothing but the breast. The caveat is the condition, and the condition is strict.

They also got things wrong, and the wandering womb, which was supposed to roam the body and press on other organs, is the one to remember.

The womb does not wander. It has, as far as anyone can tell, never been anywhere.

The household ledger

Above the medical folder sits the legal one, and it is thicker, because every civilisation that could write wrote down who was allowed to own what. Here the layer is culture throughout, and the first thing the ledger shows is that the written rule and the lived arrangement are two different columns. Chapter 2 found that a society's treatment of women is many things that do not travel together. The civilisations fill in the table.

Athens gave a woman a guardian and a set of rooms at the back of the house. Sparta let women inherit, own and run land, so that by Aristotle's time they held about 40 per cent of it and managed the estates while the men were in barracks, which Aristotle thought was ruining the place. Rome began with a marriage that passed the wife into her husband's hand and shifted to one that left her in her father's line and, once he died, her own person in law. In Han China the statutes dug up on bamboo let women head households and inherit, and preserve a case in which a Widow Wan sued her husband's male relatives for her property and won.

The lever, wherever the lived column was fuller than the written one, was property. Nowhere is this clearer than in the cities of Mamluk Egypt and Syria, in the study Yossef Rapoport published in 2005. The law gave a daughter less than a son and gave the husband the right to divorce. It also gave the bride the *mahr*, a sum paid to her and not to her father, which stayed hers throughout the marriage, which the husband forfeited in part if he divorced her, and which let her buy her own way out if she chose. The result was a divorce rate in the medieval Islamic cities approaching a third. In England, by contrast, a wife's legal person was folded into her husband's, and the only way round it was to be a *feme sole*, unmarried or a trader by town custom, until the late nineteenth century.

Here is the second thing the drawer teaches. Formal subordination rarely meant powerlessness, and wherever women held power under laws that said they should not, the lever was property.

Then the exceptions, most of which belong to the chapter before this one. Egypt coded the office of pharaoh as male, so that Hatshepsut, to

hold it, had herself carved with the false beard and the kilt. Exceptions like hers did not dissolve the categories. They drew the line somewhere else.

Why did the line get drawn where it was, and why has it moved? Chapter 2 gave the plough. The research of the last decade has given the herd. An economist traced living people across most of the world's countries to what their ancestors did for a living, and found that descent from herders predicts restricted movement for women, infibulation, and present-day opposition to abortion, the argument being that a man away with the animals could not watch his wife, and so his culture grew rules that did it for him. A team writing in *Science* argued that the Western Church's bans on cousin marriage dissolved Europe's clans and so produced the strange, individualistic West, a thesis still contested, so take it as a hypothesis.

Persistence has a model of its own. Two economists divided the centuries from late antiquity to 1900 into generations and measured how much the climate swung from one to the next, and found that where it stayed stable, so did custom, because what the grandparents did still worked. The weakening, when it came, came from two directions at once. Nineteenth-century men, who held every vote, gave women property and then political rights, and the best explanation is that once schooling paid, men wanted household money spent on the children's education, and women, on the evidence, spent it that way. And when American states began letting young unmarried women obtain the pill, the women who got it early had fewer children before their early twenties and were more often in work by the end of them. My mother was born in 1946 and I in 1975, which puts the two of us on either side of that hinge.

Note the layer, because this is the one place in the chapter where the argument crosses it. Every one of these is a cultural explanation. But the plough, the herd and the pill are all answers to a body: to the shoulder the plough wanted, to the paternity the herder could not see, to the pregnancy the pill made optional. Culture here is physiology with a long memory.

What the scriptures filed

Between the doctors and the lawyers sits a thinner folder, and it is the one most people reach for first: what the religions and philosophies said the two sexes were *for*. The doctrine, almost everywhere, is complementarity. Two halves, each incomplete, together whole. The layer is culture, though every tradition in the folder insists it is describing the body.

Genesis has God make the woman as a helper for the man, and the English word carried centuries of subordination. The Hebrew is *'ezer kenegdo*, and a Hebrew scholar pointed out in 1983 that *'ezer*, where it occurs elsewhere in the Hebrew Bible, is mostly God, coming to rescue Israel, so that the phrase reads less as "assistant" than as "a strength that stands facing him". The rabbis who inherited that text did something the Greek-minded church did not. They refused to renounce the body, and when the Talmud needed to assign religious duties it found that two sexes did not cover the people in front of it and used six words: male and female, a person with both sets of organs, whom Rabbi Meir called a creature of its own kind, a person whose sex could not be told, a woman who did not develop, a man who did not. That is a legal system filing variation in the body without pretending it was not there. The Sambia of New Guinea had a word for children born looking female who turned male at puberty, which is what a missing enzyme does, and the word was there before the enzyme was. A fact of the body, given a cultural name. The three layers, filed correctly, by people with no idea they were doing it.

The Christian line ran from Aquinas, for whom grace perfects nature and the sexes complete each other, to John Paul II, whose Theology of the Body remains the most elaborate modern statement of complementarity as a doctrine, and which its critics read as hierarchy with the manners improved. Further east the symbol is kinder and the use of it was not. The yin-yang circle puts a dot of each inside the other: no one is all one thing. A Han philosopher took the same circle and made it a floor plan, inner and outer, and assigned the wife the inside.

What can be tested is what the doctrine does when it lands in a questionnaire. Two psychologists built one in 1996 and found that sexism has two faces that travel together, a hostile one, which dislikes women who compete, and a benevolent one, which cherishes women who do not. Benevolent sexism is complementarity with the theology taken out and the

manners left in. A dot of yin in the yang is a beautiful idea. It has also been, for most of two millennia, a reason to keep the yang in the outer room.

The last unsettled fight

Now the top folder, the one still being written to, and the only part of the drawer where the three layers are fighting each other in public. Chapter 5 told one fight from it. Here is the rest of the file.

The claim, stated in the behavioural layer where it lives, is this: on a range of self-reported traits and preferences, the average gap between men and women is larger in richer and more sex-equal countries than in poorer and less equal ones. The personality studies found it first, and the values studies after them. Then the largest survey of economic preferences ever run, published in *Science* in 2018, measured risk, patience, altruism, reciprocity and trust in dozens of countries holding most of the world's people, by questionnaire and by choices with money on the table. Worldwide, women were a little less willing to take risks, a little less patient, a little more altruistic, a little more trusting. Small gaps, wide overlap, most women and most men inside each other's range in every country. But the width of the gaps tracked the country's wealth and its equality index closely, which is to say they were near zero in the poorest countries and largest in the richest. The paradox, on the preference data, is not that women and men differ much. It is that they differ most where they are freest to be the same.

The first column's reading is that when a girl in Helsinki is free to follow what she likes, she follows it, and what she likes, on average, is not engineering.

Now the other column. Thomas Breda and Clotilde Napp, in 2019, measured something no cross-country comparison had: the gap inside one pupil between her reading score and her mathematics score. A girl who is excellent at mathematics is, more often than a boy of the same mathematics, even stronger at reading. Allow for that one within-person difference and three quarters of the gap in intention to study mathematics goes away, and the gaps in mathematical confidence and declared interest disappear. Implicit association tests from dozens of countries had already shown that the strength of a nation's science-is-male reflex predicted its

science gap. Then, in 2020, Breda, with Elyès Jouini and Napp, went back to the booklets, to the very box the girl in Helsinki ticked, and found that in 71 of 72 countries girls were more likely than boys of the same ability to blame a failure on lack of talent, and that the gap was largest in the wealthy countries. The girl in Helsinki, on this reading, is not following what she likes. She is following what she has been told she is like, and the telling is louder in Finland than in Tunisia.

Then the review that should have settled it and did not. Agneta Herlitz and her colleagues went through the big studies in 2024 and found that the paradox is real and that it is particular. In richer and more equal countries the gaps grow for personality, verbal ability, episodic memory and negative emotion. They shrink for sexual behaviour, for what people want in a partner, and for mathematics. Whatever is widening the one set is not widening the other. The two large theories in the field, the one that says freedom reveals a difference in the body and the one that says prosperity breeds an essentialist culture, each explain the half of the list the other cannot.

So here is the honest sentence, and it is the sentence the Preface promised. The pattern is robust, it is behavioural, and it is cross-sectional. The mechanism is unknown.

Not disputed, not contested, not "complex": unknown, in the plain sense that the two best explanations make opposite predictions about the stereotypes, the stereotype data favour one, the preference data favour the other, and nobody has yet run the study that would decide it, which would have to follow the same people in the same country as that country changed. Anyone who tells you the paradox proves what women are really like is handing you a folder. Anyone who tells you it has been debunked is handing you a different folder. The drawer holds both.

What it hands you instead is the tool the rest of this book has been assembling. Every page in the drawer can be filed on one of three shelves. The physiological shelf, where menstruation and lactation and the Sambia child go, and where the old doctors, for all their rennet, mostly got the observations right. The behavioural shelf, where the paradox lives, as an average with an overlap and no mechanism. And the cultural shelf, where the plough and the herd and the Church and the pill go, and the floor plan, and the helper, and the girl's belief about her own talent. The civilisations never had the shelves. They had one drawer and put everything

in it and then argued about the drawer. You have the shelves. The next chapter, which is about what to build with them, calls the result Emancipation 2.0, and it begins at the scale this one ends on, the world.

THE STRONGEST CASE AGAINST

The chapter's quiet thesis, that something stable in the body shows through the variety of custom, has two serious opponents, one from each end of the drawer.

The first is the modern one. Breda, Jouini and Napp's result is not a quibble about method. It is a prediction that the first column failed. If widening gaps in rich countries were freedom revealing a disposition, the stereotypes about that disposition ought to weaken as the countries grow freer, because fewer people would need them. They strengthen. Girls in the richest countries blame their failures on talent more than equally able boys do, and the within-pupil reading advantage removes most of the intention gap before any disposition is invoked. Add the measurement objections. That computing a sex difference as a score and then correlating that score with equality overstates the relation by construction, so that part of the paradox is arithmetic. That a woman in a tight society rates her own assertiveness against other women and a woman in a loose one against everyone, so the questionnaire is measuring the comparison group, not the trait. And the test chapter 5 tells, in which a country growing more equal over time does not see its own gaps widen. Put together: the gaps may be real, and the paradox a fact about which countries are rich, which is a fact about history, not hormones.

The second opponent is Laqueur, and he survives his own defeat in a weaker form. The words for the body were still, always, cultural, and the chapter's own medical folder shows how large the cultural term can be: the same premenstrual syndrome, rare in France and near universal in Iran. If a filter can swing a measure of the body that far, the historian is entitled to ask how much of any cross-cultural sex difference is the thing and how much is the filter, and the chapter cannot answer that for the paradox. Laqueur's critics could not answer it for the skeleton either.

I think the first opponent is winning on the science data and losing on the preference data, and I think the second has a point that no amount of

data can dismiss. I do not know who is right. I know that the fight is on the behavioural shelf, and that both sides keep reaching for the other two.

IN PRACTICE

The world is the one scale where you cannot run the experiment, so what follows is about reading, not doing. These are conditionals for the person who meets a cross-cultural claim about the sexes, which is to say everyone who reads a newspaper.

If you read that a sex difference is larger in Sweden than in Saudi Arabia, the first useful question is which shelf the claim sits on. If it is a questionnaire about what people say they prefer, it is behavioural, cross-sectional and open to every objection in the section above, and the sentence that follows it, about what women really want, is not licensed by it. If it is a measured outcome, strength or survival in a famine, it is physiological and the cultural reading has to work harder. The failure mode is the reverse: reading a questionnaire as if it were a blood test, or a blood test as if it were a preference.

If you are tempted by the freedom-reveals story, because it flatters a side you are on, the test is to ask what it predicts about stereotypes, and then to look. If stereotypes weaken where gaps widen, the story holds. If they strengthen, as Breda found, the story has a hole in it. If you are tempted by the stereotype story for the same reason, the test is the other half of Herlitz's list: a mechanism that explains why girls in rich countries avoid engineering has also to explain why those same countries show larger gaps in episodic memory and negative emotion, which no one is stereotyping anyone about. The failure mode on both sides is the same. It is keeping the half of the list that agrees with you.

If you hire, teach or manage across cultures, and you have in your head a map of which countries produce which kind of woman or man, the map is likely to be a map of averages measured in one year, and the person in front of you is a point, not an average. In any country, most women and most men sit inside each other's range on every preference the big survey measured. The failure mode is treating a country's bell as a person's label. If you correct for the first by assuming no differences at all, you are likely to build a workplace that fits the people who already fill it and then be surprised that it does not fill with the others.

If you inherit a norm, a floor plan, a rule about who sits where, look for the body it answered. The plough answered a shoulder. The herd answered a husband's absence. If that body is no longer in the room, because the plough is a tractor and the pregnancy is optional, the norm is running on memory, which it tends to do for as long as nothing disturbs it. The failure mode is assuming that a norm with a reason once must still have one, or that a norm without one now never had one. Both lead you to push on the wrong wall.

If you are the one trying to change a norm, the ledger has one more thing to say. In every civilisation where the lived column was fuller than the written one, the lever was property. The *mahr*, the Spartan estate, the trading licence. The men of the nineteenth century gave rights when it paid them to. If a change you want does not change who holds what, it is likely to be a change in the written column, which is the column history has shown it can leave without effect. The failure mode is believing a declaration.

If you want to use the paradox as evidence for anything, the evidence it supports is narrow. It supports the sentence "cross-sectional sex differences in self-reported traits are larger in richer countries". It supports nothing about why. If someone uses it to tell you what to do about girls in science, or boys in nursing, they are using a correlation with no mechanism as a mechanism. The failure mode is the one the Preface warned of: finding the drawer, pulling out your folder, and calling it the drawer.

And if you read the old doctors and the old lawgivers and are tempted to laugh, the lesson of the medical folder is that they saw most of what we see, and filed it wrongly, and that the filing was the part culture did. The failure mode is assuming that the filing we do now is not filing. The girl in Helsinki ticked a box. Somebody chose the box.

The studies, figures and open questions behind this chapter are on the website.

CHAPTER 16

Emancipation 2.0

It is a quarter to seven on a weekday morning in a flat in a German city, and two people are standing in a kitchen with one cup of coffee between them. One of them is about to leave for the day. The other is about to take a small child to a nursery of a kind that, in the half of the country I was born in, was a rumour. Neither has said anything yet. The kettle has. I am not going to tell you which of the two is leaving, and I would like you to notice, before you read on, what your mind just did with that.

In 1971 it needed no holding, because there was one answer, and the language had an insult ready for the woman who gave the other one. Today there are two answers. The tax form on the shelf behind them has had a view on the matter since 1958. None of those views was arrived at by asking the two people what they wanted.

Everything in the book is in this room. The hands around the cup are physiology, measurable. The two people are statistics, two draws from two bells that mostly overlap, so that whatever you guessed about who is leaving, you were guessing at a population. The form, the timetable and the insult are culture, variable: a wall down the middle of this country moved one of those bells and left the other where it stood.

My mother's generation asked one question of that kitchen and answered it so loudly that the sentence seemed finished. *Can she?* She can. The sentence had a comma after it, and nothing was ever written after the comma. The second half, how would she and he want to, could not be asked without sounding like the people who had wanted the old answer.

So it went into the drawer.

This chapter is the rest of the sentence. It contains no new fact. Stay in the kitchen. It is the only laboratory the book has left.

What the first half bought

The first half of the question has a receipt, and it is worth reading before the second half is asked, because it is the strongest thing in the book.

A woman can do anything a man can do, in the sense that mattered in 1971, and in the eastern half of this country she did it while raising the child and out-earning her husband. Barbara Boelmann, Anna Raute and Uta Schönberg took a huge random sample of German social-security records and found that a generation after the wall came down an East German mother still works about 350 hours a year more than a West German one, and that most of the gap is the stigma and none of it the wage. Chapter 13 has the whole of that story. Give a population a tax code, a timetable and a word for a generation and you move its mothers, and the movement outlives the regime and travels in a suitcase. Culture is the largest effect in this book.

Here is the rest of the receipt. The wall moved the mothers. It did not move the fathers, in either direction. The East abolished the breadwinner norm and kept the same woman doing most of the laundry and most of the thinking about the laundry, and in reunified Germany's family panel couples who start symmetrical still tip at the first birth, East and West alike, and do not tip back. Rebecca Sear and Ruth Mace went through every population in which anyone had kept count of which relatives were alive when a child lived or died, and found the mother's mother improving a child's survival in about two in three of them and the father in about one in three. And in the richest and most equal countries the gaps in interest, risk and personality did not close. They widened, and after chapter 15 I still cannot tell you why.

That is the receipt, and it has two columns, and my mother's generation read only one of them, which is the column I would have read too. They had an abortion law to fight and a husband's signature to abolish, and a question about what she and he would *want*, asked in 1971, sounded exactly like the question their opponents were asking. Chapter 5 gave the

reflex its name: if a difference would be unfair were it true, it cannot be true. The moralistic fallacy is the fallacy of people who have been lied to about nature before and have decided, reasonably, not to be lied to again.

So why ask the second half now? Because the kitchen is already designed. The income splitting of 1958 is a design, drawn from an assumption about what a wife is for. The half-day school is a design. The fenced pitch, the toilet floor plan, the long week, the sentence that runs longer for the same file, the standard written for one skeleton: every one of those is an arrangement somebody chose, and most were drawn either from a difference nobody had measured or from a sameness nobody had checked.

There is no item on the menu called no design. There is design from folklore, design from denial, and design from the evidence, and the first two have had their turn.

Seven scales, then, one for each the book has climbed, and at each one a hypothesis with the evidence against it on the same line, in the grammar the Preface could only promise: given all three layers, what arrangement tends to produce what.

The couple, the family, the team

The couple. The most quoted finding at this scale is the one chapter 6 watched fall: a ratio of good exchanges to bad, and a prophecy of divorce read off a few minutes of tape. Tested on couples it had not been fitted to, the prophecy lost most of its power, and the ratio itself traces back to a couples' app. The folklore's principle is a risk factor dressed as a ratio, and I do not carry it.

Flooding is measurable: past a threshold of pulse the body stops listening whatever the mouth is still doing. Who floods is statistical, and depends on the instrument. On the heart-rate monitor men flood sooner. On the adrenal glands women rise higher and for longer, most of all when the husband withdraws. And who demands and who withdraws is culture. The demander is whoever wants the change, not whoever is the woman.

So the principle: given that one body in the argument tends to leave before the person does, a couple that agrees the stop rule before the first flood, and agrees who is asking for change this time rather than

who is the woman, tends to keep the argument from becoming the pattern. When thousands of couples were run through machine learning, everything about the two individuals combined, sex included, explained a small part of how good the relationship was, and the strongest predictor was whether a person believed their partner understood them. If the accent is in the couple at all, it is a quiet one.

The family. The second principle is the one this book was built on: a family holds together when provisioning and care are spread across people who are not the child's parents. Sear and Mace's count carries most weight in the book, the grandmother in two in three populations, the father in one in three, and chapter 9 brought it indoors, to the woman who knows where the plasters are kept and appears in no statistic the city collects.

Eyal Abraham, Ruth Feldman and their colleagues put first-time parents into a scanner: mothers, fathers whose partner did most of the care, and gay fathers raising an infant with no mother in the house. The brain did not ask who had been pregnant. It asked who was up in the night, and across the fathers the connection between the old circuit and the new rose in a straight line with his hours of care. The hormone fell with the hours too, as chapter 8 told from Cebu. The father's share in the survival count is small because, where survival was counted, he was rarely the one holding the child. Where he holds it, the body follows.

So the principle: given that no human child has ever been raised by two adults alone for long, a family that builds the third adult in, by kin where the kin are near and by paid care where they are not, tends to do what the pod does, and given that the parental body follows the caregiving hours and not the sex, the parent who takes the hours tends to become the parent the data describe. The counter-evidence is in three places. Chapter 9's question about the stress hormone that rises through the afternoon at nursery is still open. The parish books of Krummhörn, where a paternal grandmother in the house raised the chance that a newborn died, say that "kin" is not one word and the third adult has to be chosen, not assumed. And the father's one in three is not a ceiling anyone has tested in a population where fathers do half the hours.

The team. The third principle is the one whose factor fell. Chapter 10 watched a group intelligence that rose with the number of women at the table come apart on second look. What the second look left standing was a pair of findings about doors and rooms rather than about minds. The

room: in teams working by voice alone, men said almost 70 per cent more words than women, and the ablest women said least. The door: when a field trial put a job advertisement in front of thousands of job seekers and made competitive pay something you had to tick a box to get, a gap of nearly 26 points opened between the men and the women who applied. Make the competitive pay what you got unless you ticked the box to refuse it, and the gap was a single point. Nobody's preference changed. The form did.

So the principle: given that the measurable sex gap in a team sits at the door and in the airtime and not in the work, a team that changes the form and names the chair tends to close more of its gap than a team that changes its headcount. The counter-evidence cuts both ways. From one shelf, the economists who found that the entry gap may be risk and confidence measured with error, so that the preference the form seems to override may never have been a preference at all. From the other, Alice Eagly and Wendy Wood's finding that most measured sex differences mirror what happens when you hand one group power over another. If airtime tracks power, then naming the chair fixes power, not sex. The two people in the kitchen do not need to know why the form works to use it.

The organisation, the city, the state, the world

The organisation. The design the folklore offers here is to tie advancement to unbroken hours, which selects for the sex with the looser claims on its evenings. Chapter 11 found the long week to be a measured killer, and it kills mostly men: in one year about three-quarters of a million deaths from heart disease and stroke were laid at the door of the long week, and most of the dead were men. And the labelled accommodation is used by nobody and paid for by the people it names. Japan's menstrual leave sits almost untouched, and a benefit mandated for one sex comes back as lower wages for that sex.

So the principle: given that the unbroken week damages mostly men's bodies and the labelled leave prices mostly women's hours, an organisation that pays for the outcome rather than the hour, builds the accommodation nobody has to be named to use, and writes down who answers to whom, tends to keep more of both sexes than one that does any one

of the three alone. The counter-evidence is the book's own. Where the task is the body, no reward structure moves it: in the Marine Corps experiment the all-male teams came out ahead in most of the tasks, the stretcher among them. Decoupling reward from the hour is a design for the desk, not the stretcher.

The city. Build home, work and childcare within reach of each other, the folklore says, and the divide loses its friction. Chapter 12 found the city carried by one city's account of itself and said so. In Montréal a quarter of men's trips are care trips, so the loop is the shape of whoever does the care and not of a sex. A woman takes longer at the toilet than a man, so equal floor space produces a queue by arithmetic. And Vienna, where the planner Eva Kail began the work in 1991, has been the one city that counted the loop before it built for it.

So the principle: given that the care loop in every city measured is walked mostly by women today, a city that counts the loop and builds for the loop, not for the women, tends to shorten the day of whoever walks it tomorrow. The counter-evidence is the plough. The tool a society farmed with centuries ago still predicts who leaves the house today, and a city that makes one arrangement comfortable tends to keep it, so the loop you build for may be the loop you fix in place. Proximity priced by a market becomes a luxury. And chapter 12's hardest question stands: whether the city is about sex at all, or about care, which any adult can do.

The state. The principle the Preface promised: the best single predictor of a state's peacefulness is the physical security of its women, a finding chapter 14 carries and the critics have fought over since. The same folklore adds that legal polygyny breeds a surplus of unmarried young men and instability, and here the book has a hole: the research commissioned for chapter 14 never arrived, and I leave the clause open rather than fill it from memory.

What chapter 14 did establish is why the principle must be read as a fact about the state and not as a promise about the women. Oeindrila Dube and S. P. Harish went through four centuries of European reigns and found a state under a queen 39 percentage points more likely to be at war in a given year than a state under a king, because a married queen could hand her husband the army and keep the state, and a married king did not. Women in Indian state legislatures cut newborn deaths through antenatal visits and iron, not through gentleness. And under identical Is-

raeli training the women's bones broke at a multiple of the men's rate.

So the principle: given that a state's peace tracks the safety of its women and not their supposed gentleness, a state that designs for the safety tends to get the peace, and a state that designs for the gentleness tends to get the queens. The counter-evidence is in the direction of the arrow: a state which violates its women may be unstable for a third reason that produces both. On the standards the counter cuts both ways. A sex-blind standard breaks women's bones, and a sex-lowered standard breaks the trust of the men beside them, so the design is a standard set to the task and not to the sex, with the training built to the bone.

The world. The last principle in the folklore, that settlements last when women hold real negotiating power, rests on a citation I could not make stand up, so it goes to the website and not into this paragraph. What the world scale hands the chapter is the finding I like least and trust most. Across dozens of countries the gaps between the sexes in risk, patience and personality are widest in the richest and most equal places. The best counter-study finds the stereotype that mathematics is not for girls *stronger* in exactly those countries, and the paradox gone once it is allowed for. And the one study that watched countries change, rather than comparing them, found no gap widening inside any country as it grew more equal. Three good studies, two shelves, no verdict.

So the principle: given that the measured gaps have widened, not closed, in the countries that opened the most doors, and that nobody can yet say whether that is freedom or a subtler stereotype, a world that designs for equality of the door tends to get more difference at the table, not less, and should budget for the difference rather than read it as the door having failed. The counter-evidence is the whole of the facing shelf, and the honest sentence is the one chapter 15 ends on. I don't know. Neither does anyone else yet.

What she would recognise

My mother was born in 1946 and died in 2021, and in between she watched the first half of the question get answered. I think she would recognise more of the second half than she expected to.

She would recognise that every principle above begins where hers did.

The state's peace begins with a woman's safety. The organisation's design begins with a woman's hour not being priced. The team's begins with a form that does not quietly enrol him and ask her. The city's begins with counting the trips she makes. The first half was not wrong. It was a prerequisite. You cannot ask two people what they want from a kitchen until both of them are allowed to leave it.

She would not recognise three things. The first is that the father is in the design, not as a second mother but as a body that changes with the hours it does. The second is the grandmother, who had been in the design since before anyone could write and was removed from it, in the West, by a tax form and a half-day school that assumed a mother at home and no one else. The third is that the second half of the question has no arrangement as its answer. It has a grammar. Given what we are, under these conditions, this tends to produce that. It does not say what she or he should want. It says what tends to happen when they want it, and it leaves the wanting to them.

That is Emancipation 2.0, and it is not a second wave. It is the rest of the sentence her generation started, written in the only grammar that survived the border.

So the tool. At every scale the book has climbed, the same four questions, in order. What is measurable here, and in which body? What is statistical here, and how far do the bells overlap? What is culture here, and who drew it, and when? And then, only then: given all three, what arrangement tends to produce what, for whom, and what is the evidence against it?

Ask them of the kitchen. Then ask them of the two people in it, because the two people are not the distribution. They are one draw each, and the only thing the whole book can tell them for certain is that nobody else can tell them what to want.

THE STRONGEST CASE AGAINST

The case is against designing anything from these differences at all. Its best voice is Alice Eagly and Wendy Wood's, and it goes like this.

After correcting for the larger male head, sex explains almost none of the variation in how brains are built. Adult hormone levels are not a

cause waiting to be designed around but a readout of the life being lived: care lowers testosterone, authority raises it. Most of the sex differences the meta-analyses find mirror the effects produced when experimenters hand one group power over another. Caregivers lay gender into infants in their first months. And the wall: most of the German mothers' gap is the stigma, and none of it is the wage.

It is that every design drawn from a difference manufactures the difference it was drawn from. The income splitting of 1958 was drawn from an assumption about wives and produced, two generations later, wives who fit it. If culture becomes biology in the first year of life, then a fixture ratio, a care loop built into a bus network, a crèche sited for the person who does the care, are stereotypes poured in concrete, and the person who walks the loop tomorrow finds it already waiting. Anything else is the naturalistic fallacy with a planning permit.

That case is right about more of this chapter than I would like. The couple's stop rule, the team's form, the organisation's unlabelled accommodation, the city's loop, the state's safety: those are culture-layer designs, and every one of them would be worth building if the Endocrine Accent were silent. Only three things in the chapter lean on the body: the gestation and lactation no agreement moves, the bone under the pack, and the stretcher.

Where I still part from it is the kitchen in the East. The constructivist case predicts that where the breadwinner norm was abolished for the whole life of a state and the insult ran the other way, the laundry should have moved. It did not. The fathers did not move in either Germany, under either insult. Something in that stillness is either a second wall nobody has found or an accent nobody has measured, and a design that assumes nothing about it is still a design, drawn from a sameness nobody has checked. The case against has the stronger numbers. It does not yet have that kitchen.

IN PRACTICE

The scale is yours, whichever it is. One conditional for each, with the way it fails and the evidence against it.

If you are one of the two people in the kitchen, and you are about to decide who leaves at seven, then write the decision in the book's grammar be-

fore the first birth and not after it: given our two jobs, our two bodies and the hours the nursery keeps, this arrangement, until this date, and we look at it again then. The failure mode is the one chapter 7 found in the panel: the decision made in the pregnancy tends to be overruled in the third month of the baby by lactation, which is the one bottleneck no agreement moves. The counter-evidence is the eastern half of the same panel, where the same bodies tipped differently, which tells you that everything after the milk is negotiable.

If you are building a family, then build the third adult in before the first child arrives, by kin where the kin are near and by paid care where they are not, and if it is kin, choose which. The failure mode is Krummhörn: "family" is not a design, and a named person is. The counter-evidence is chapter 9's open question about the cost of paid care, still open.

If you run a team and you want the gap closed, then change the form and name the chair before you change the headcount. The failure mode is that the form fixes the door and not the room, and in the room the men tend to do most of the talking. The counter-evidence is that the gap at the door may be risk and confidence measured badly rather than a preference at all.

If you run an organisation, then pay for the outcome and not the hour, and build the accommodation nobody has to be named to use, for everyone, including the man with the migraine. The failure mode is the senior man's own week: a policy tends to be the hours the most senior man is seen to keep, and the long week is a measured killer of men before it is anything else. The counter-evidence is the stretcher, where no reward structure moves a load.

If you sit on a council, then count the care loop before you change a route, and build for the loop rather than for the women who walk it today. The failure mode is the plough: a city that makes one arrangement comfortable tends to keep it. The counter-evidence is one survey in Montréal and one city's account of itself.

If you vote, draft or judge in a state, then design for the safety of its women and not for their gentleness, and set every physical standard to the task and the training to the bone. The failure mode is the cabinet that appoints a woman to the war ministry and expects peace, with the queens as the record. The counter-evidence is the arrow: a state that violates its women may be unstable for a third reason that produces both.

If you are designing for the world, then budget for difference under equality rather than reading it as the door having failed. The failure mode is the verdict: reading the paradox as proof that the accent is loud, when the best counter-study finds the stereotype louder in exactly those countries, and the within-country test finds no widening at all.

If you catch yourself using an average on the person across the table, then stop. The bells overlap and the person is one draw. The same overlap, used to deny the average when planning a city, is the other jaw of the trap.

And if you are reading this book, then apply its method to it. Every figure in it stands on a named study that is still being checked against the paper, and the checking is on the website. The failure mode is mine: the narrator has a side, a sex and a generation, and a book about men and women that cannot lose an argument to both tends to be read as a prosecution. The counter-evidence is the facing shelf, which has stronger numbers than my side on the brain, the hormones and the negotiation.

And if none of this fits the two of you in the kitchen, then the overlap is doing its job, and the question my mother's generation asked is yours to finish. Not whether she can. That half is done. *How would she and he want to.* It is a quarter to seven. The kettle has already answered. The two of you have not.

The studies, figures and open questions behind this chapter are on the website.