

Principles of Psychology and Neuroscience, First
Edition

Randall C. O'Reilly

Principles of Psychology and Neuroscience

(The Three C's)

First Edition

Randall C. O'Reilly



Contents

Preface	4
Chapter 0: Introduction	5
The Three C's	6
Compression	6
Contrast	7
Control	10
The Breakdown of Control	12
Other Principles and Perspectives	13
Where do we go from here?	13
Chapter 1: Science and Subjectivity – The Fundamental Challenge of Psychology	15
Subjectivity in Psychology: A Brief History	17
Fundamentals of Cognitive Neuroscience	18
Subjectivity and Science: Working with the Method	19
Research Methods in Psychology and Neuroscience	21
Neuroscience methods	24
Statistics	25
Chapter 2: Neuroscience	29
Simple Neurons Make Complex Work	31
The Tug-of-War in Your Brain	34
Large-Scale Brain Organization (“Gross” Anatomy)	37
Acknowledgments	39
Glossary	40
About the Authors	41
References	42

Published by Open Textbook, freely available

Copyright © 2018 Randall C. O'Reilly

All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law. For permission requests, write to the author, addressed “Attention: Book Permissions,” at the address available from below.

<https://github.com/PsychNeuro/ed1>

To my family.

Preface

This is an in-progress experiment – feedback is more than welcome.

Chapter 0: Introduction

Introductory Psychology textbooks typically provide a rather fragmented, fact-laden view of the field, relying on colorful graphics, exciting news stories, and personal anecdotes to generate interest in the material. This book represents a radical departure from that approach. Instead, the goal here is to provide a simple, succinct, principled account of the human mental world and how it emerges from our brains, that is coherent across the scope of phenomena covered in typical Intro Psych texts.

The overall portrait painted of you looks something like this (caution: it is not overly flattering, and you might even think this song is not about you, but go ahead and be vain – it is): You are obsessed with controlling your environment to satisfy a range of core desires and to mitigate strong fears. You are unlikely to be swayed by other people's advice, but have no problem dishing it out. A challenge to your social standing or any other form of disrespect (the *diss*) is one of the worst offenses. You are willing to spin all manner of stories to maintain your sense of order in the world, *especially* when that sense is strongly challenged, often to the point of absurdity in the eyes of others.

You crave simple ways of understanding the world, to the point of massively oversimplifying the true complexities and ambiguities, preferring to think in terms of concrete anecdotes instead of broad abstractions, logical arguments, or, especially, statistics. You think you know how most stuff you use everyday works (bikes, cars, toilets..), but studies show that you are actually remarkably clueless – how exactly does that chain on a bike work? Perhaps most glaringly, you can't help but think in terms of stereotypes, and inevitably focus on information that is consistent with your existing views, while ignoring all those nagging hints that all may not be as simple as you might like.

You only care about things that are new and unexpected, and are constantly comparing and evaluating yourself and others with a keen eye for who is doing better or worse along any number of important dimensions (wealth, beauty, smarts, athletic ability, popularity – you name it!) You are hypersensitive to who might be cheating or gaming the system, but are perhaps not so aware of unfair advantages you might have. More generally, you tend to think of yourself as being “your own person” and strongly underestimate how strong of an influence other people actually have over you. If you're honest with yourself, you'll admit that you spend way too much time thinking about what other people think of you – without recognizing that everyone else is doing the same thing, so that in fact the answer is a somewhat disappointing: “not much” (unless of course you do something embarrassing or strange or stupid, but even then, your memory of those events will typically far outlast those of others).

In other words, you are a *survivor*. You are a tough cookie. Your ancestors survived unbelievable hardships to get you here, to your relatively plush college-educated world. You are amazingly efficient. All those crazy details you don't know about the world are largely irrelevant anyway. Seriously, does it really matter that you don't know how the engine or transmission in your car works?

You can drive, and get to where you need to go – and that is what really matters. Your brain is exquisitely tuned into what really matters, and despite over 60 years of attempts to recreate the magic of your brain in a computer, nothing has come even close (despite all the recent media hype to the contrary).

And yet, despite all your toughness and amazing abilities, you are very likely to have at least some level of significant mental dysfunction. You are more likely than not to suffer from depression, anxiety disorders (and often both of those together), drug dependence, etc. Unfortunately, the promise of a magic pill to cure these afflictions has turned out to be yet another disappointment. In fact, regular old “talking to another human being about your problems” (i.e., therapy, which is actually somewhat more involved and structured than that) is likely to be more effective than medication for most people.

The Three C's

Surprisingly, we can make sense of all the above (and more!) using only three core principles:

Compression

Each neuron in the most important part of your brain (the *neocortex*) is wired for simplification, and the collective effect of the massive waves of electrical activity surging through your brain every millisecond is to compress, reduce, and simplify information. Each neuron receives input signals from roughly 10,000 or more other neurons, but guess how much it can then say about that flood of information coming in? Almost nothing. First of all, it only has *one* output signal, the *spike*, which is an all-or-nothing affair. Furthermore, a typical neocortical pyramidal neuron will fire at most around 100 spikes in a second. And a second is a relatively long time in the inner loops of the brain – there is evidence that 1/10th of a second represents a kind of fundamental time-frame for information processing, so those 100 spikes reduce down to just 10 spikes within that critical window. And most neurons are firing far less rapidly than that. It's like when you tell your friend all your deepest thoughts, and they just say “huh”. Neurons are the strong, silent type most of the time. But still waters run deep: when neurons *do* get excited about something, it is likely to be *important*, and most of what they are doing is *shielding you from constant TMI* (too-much-information – but you knew that already, so, kind of a meta thing we got going there...)

The raw scene coming into your eyeballs is truly gory: all jumbles of light, motion and color. When you were a tiny baby, you were overwhelmed by this “blooming buzzing confusion”, but now your neural networks have learned and developed to the point where you don't (can't!) even see that raw sensation anymore (unless of course you partake of various hallucinogenic substances, but even then, the level of disorder experienced is trifling compared to the pure chaos of the raw, unfiltered tidal wave of sensation coming in). We get small,

fascinating hints of the magic power of our perceptual systems through illusions, and the occasional “viral gold / blue / brown dress” controversy, where people strongly see or hear very different things from the very same stimulus. But overall, we really have absolutely no idea how much undercover cleanup work is going on inside our brains. If anyone was truly aware of the level of conspiracy operating in there, it would be scandalous. But, somehow, amazingly, we largely all end up converging on the same stable, boring illusions of simplicity. A table. A chair. Some french fries. People walking down the street. Cars driving by. Nothing strange going on here.

We would be utterly nonfunctional without this compression. For the same reason that those hallucinogenic drugs render people nonfunctional. If you want to do something useful with your time, you need to be able to make everything else in the world boring and irrelevant, so you can focus on *what matters*. If you’re reading a book, or your tiny screen, it simply wouldn’t work if every time you moved your eyes, the whole world was seen afresh, requiring you to reorient and rediscover what you were just reading and what you need to read next. Interestingly, this capacity for perceiving a stable, boring world seems to depend critically on a very active underlying process of *prediction* – your brain is stitching everything together in a seamless whole by filling in the gaps with what you *expect* or *predict* to see. You can easily see this, and relive some of your earliest experiences, by simply closing one eye, and then gently pushing on the bottom of the eyelid of your other, open eye. Suddenly, the world is a moving jumbly mess again! (Seriously, try it!)

Your brain’s penchant for simplification (compression) does not stop with perception. Your highest levels of thought are similarly dominated by the same quest to render everything simple and predictable. Instead of recognizing the incredible high-dimensional diversity of our fellow beings, we inevitably reduce everyone to stereotypes. Even members of negatively stereotyped groups are caught in the evil maw of this process, exhibiting similar levels of stereotype-driven biases as everyone else. The ultimate expression of this compression process is the *anosognosia of everyday life* (aka the Dunning-Kruger effect; NY Times Article: <https://opinionator.blogs.nytimes.com/2010/06/20/the-anosognosics-dilemma-1/>) – the lack of knowledge about our utter lack of knowledge. People can be remarkably unaware about what they don’t know, and sometimes, this leads to funny situations. But, amazingly, most of the time, *it causes no obvious problems whatsoever*. We just keep getting on with our lives. And, as with perception, if we didn’t, we’d never get anything done, because there is such a huge amount of stuff we routinely, safely ignore, that it would take many many lifetimes to process and understand it all.

Contrast

The next principle explains why we seem so fixated on comparing ourselves with others. Not just any others, but those certain people *who really get to you*. In that inexplicable, frustrating way. Why do I always have to be so jealous of those

people? Can't I convince myself that the "grass is always greener?" Nope. As with compression, your brain is wired at the lowest level for magnifying contrasts, in this case via a special class of neurons called *inhibitory interneurons*, coupled with other important properties of all neurons that we'll cover in Chapter 2. The net effect is that your brain only sees things *relatively* (yep, we can have our own, special, relativity law in Psychology too – actually it is pretty general). A classic example of this is when you come in from the bright sunny outdoors into a dimly-lit room. The difference in raw light energy coming into your eyeballs in these two situations is enormous, but, after a brief period of adaptation, you're seeing things in the dim room that differ by a few photons here or there, whereas outside those few photons would be a minuscule drop in the bucket. In other words, our neurons *normalize* away the raw strength of whatever signal is coming into them, and remain sensitive to the *relative differences* compared to that overall signal. Those inhibitory neurons play the critical role of mathematically *dividing away* the raw signal strength, leaving the principal pyramidal neurons "in the zone" for responding to relative differences.

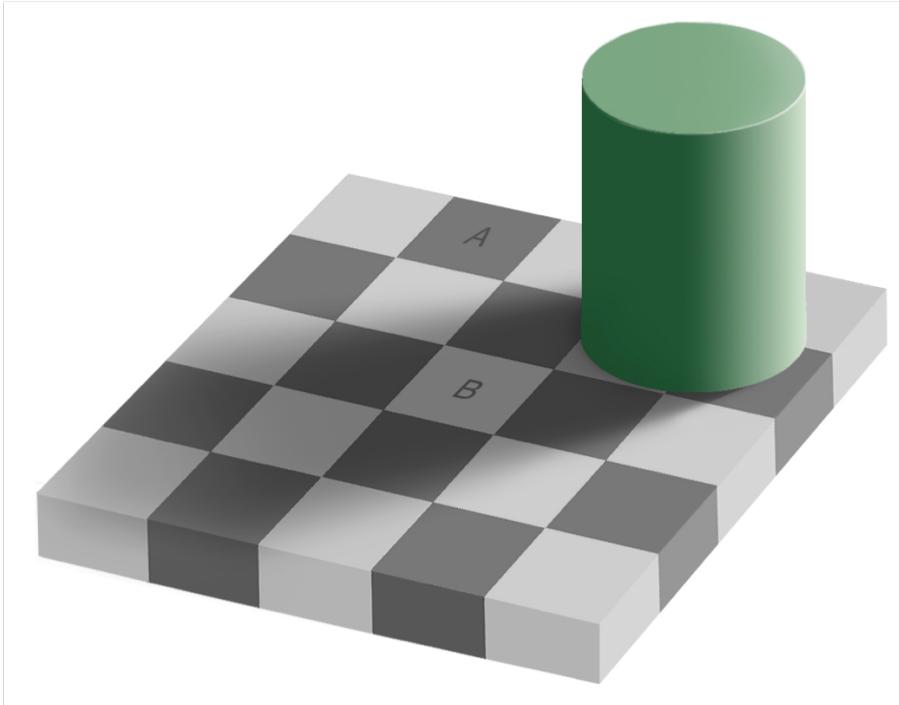


Figure 1: Fig 1-1: Illustration of the power of contrast in perception. Do you think the physical image-level color of square A is the same or different from that of B? Unbelievably, they are identical!

As with compression, perception provides some of the clearest windows into this phenomenon, for example Figure 1-1, showing the remarkable effects of

contrast (and global scene understanding) on perception of color and brightness. Another remarkable example is the case of *perfect pitch* – why is it so unusual a skill for people to simply be able to recognize the absolute frequency of a sound? Mechanically, and mathematically, extracting such a frequency from a sound signal is trivial, and simple (a “Snark”-like guitar tuning device can be had for a few bucks). That this feat is so incredibly rare and difficult in humans just points to the pervasive power of the contrast relativity (most people can easily tell the relative pitch).

But contrast, like compression, is not restricted to the perceptual domain. It affects every level of thought, contributing to that insidious obsession with your relative standing among your peers. For example, studies routinely show that the absolute amount of money that people make is largely unrelated to various measures of their happiness – instead, what matters is their perceived level of income *relative to their peers*.

Contrast operates over time as well, in several important ways. First, at the perceptual level, we are highly sensitive to the rate of change over time of stimuli. The classic example here is the slow approach to boil being unnoticed by a hapless frog until it is too late (this is not exactly true – you can’t get all the way to boiling, but it is very likely true to at least some extent). Similarly, a cottage industry of amazing demonstrations of our inability to detect slow changes in visual scenes sprung up a few years ago: YouTube Link to Dan Simons Video: <https://www.youtube.com/watch?v=1nL5ulsWMYc>. Once you become aware (upon repeated viewing or instruction) of the nature of these changes, it is truly astounding to realize how much you overlooked them the first time(s). If you rapidly flip between the start and end frames of these slow-moving videos, the changes pop immediately into view. Again, we see the *delta*, not the absolute value of things.

Nowhere is this more poignant, and pressing, than deep inside the *dopamine* system in the middle of your brain. As you already know (and would be annoyed to have me repeat, but I’m going to do it anyway, to prove the pertinent point), dopamine is widely believed to be the “pleasure drug” in your brain. It is associated with drugs of addiction, and actually most other major mental disorders in one way or another. However, this popular description of dopamine leaves out one of the most important points: dopamine is *not* about *raw* pleasure, but rather, about the *difference* between what you experience and what you *expected* to experience. Specifically, if you get exactly what you expected, your dopamine goes “meh”. This soul-crushing response to your greatest accomplishments is exactly what critics do to performers, and indeed the dopamine system is best understood as being the *central critic* of your brain. Far from a center of epicurian delights, it is a hard-nosed bully that is never satisfied. And that dissatisfaction is what has driven us ever upward in all manner of exploits – many of them good, but many of them not so good.

Greed is really a byproduct of your dopamine system. Seriously, why in the world can’t someone who already has *millions of dollars* just be happy with that, and give the rest away or do something else useful with it (and their all-too-brief

lives). Because dopamine adapts quickly to that million-dollar feeling, and it keeps giving you back that critical “meh” response. You need more than that. You *deserve* more than that... It really is tough living with such an asshole critic in your head all the time. But then again, we really do owe every step of “progress”, within our individual lives and as a society and a species, to that nasty little critic driving us ever upward and onward.

Finally, another obvious manifestation of the contrast principle is our collective obsession with the *news*. Especially with the advent of the 24hr news cycle and the constant updating of news information via electronic, online media, we are now living in a quickly-moving bubble of news that sweeps things up in its path and spits them out quicker than... yesterday’s news. Or yesterhour’s news. If you don’t check in quickly enough, you’ll miss huge swaths of news that everyone would have definitely been aware of before. Everyone worries about this kind of thing, but really it is just what our brains are wired to do. Every conscious moment of our lives is driven by a thirst for knowing what has changed, what is different – anything that remains constant will quickly drift out of your mind, like that delicious aroma of dinner that I can no longer access, or, thankfully, that feeling of my butt sitting on this chair that I was thoroughly *so over with* until I just wrote that sentence..

Control

Last but certainly not least, is our obsession with control. Some of you may be thinking that you’re not a control freak like those *other people*, but actually, every one of us is a crazy control freak at some level – it just differs in terms of what matters to us. Anyone want to have some stranger come pick you up and take you around to work with them all day? Or just invade your personal space? How would you feel if someone just started selling all your stuff on craigslist? Or how about those people who go door-to-door (or stop you on the street) and try to convince you to believe in some particular brand of religion? Or just your roommate who keeps nagging you about the dishes, or being to loud, etc. Yeah, there’s definitely *something* for *everyone*, where it matters. And usually, if you have two or more people living together, you quickly become aware of all that stuff that you didn’t realize really matters to you. A lot.

Starting again in the brain, virtually every neuron in the brain is serving the master of control at one level or another. At the most basic level, control is about *motor control*, and a great example of the dedication of the brain to this particular function comes from a lowly sea squirt that starts off life as a mobile tadpole, and flits around in the ocean for a bit, looking for a good place to settle down. As soon as it finds its special place on the reef, it promptly eats its own brain! Because, the whole point of the brain in the first place, evolutionarily speaking, is to process sensory inputs *in the service of producing useful motor outputs to improve survival and the overall quality of life*. There’s a reason nobody thinks highly of layabouts and 30-year-old’s living in their parent’s basement: progress requires action, and our brains are wired for action.

In the brains of most species, there are big chunks devoted to the compression and contrast processing of sensory inputs, and the rest is devoted to using that information to figure out what kinds of opportunities and threats are out there in the world, and how to best optimize chances of survival within the repertoire of available motor actions. Not much space left over for cultivating expertise in civil war battles, or fantasy role-playing games, or whatever other weird, seemingly non-functional things people spend their time doing.

The human brain takes this obsession with motor control to the next level, by building an internal fortress / castle of the *self*. We're not quite sure to what extent any other beast even has a similar kind of thing inside their own mental worlds. The self is a model, a construct, built up over years, that helps us predict (again) how we are going to behave, and what we seem to really want (and not want). By having such a thing inside our own brains, we can use it to more accurately anticipate what kinds of motor actions are really going to get us what we want. This is especially important when dealing with other people, who are, compared to your average rock or tree, very complicated and unpredictable. I'm not saying you're a manipulative little jerk. I'm saying *everyone* is a manipulative little jerk, deep down. It is, again, just a logical extension of what brains are supposed to be doing. If they aren't good for maximizing pleasure and minimizing harm, then we might as well all just eat them for dinner!

This *self model* lying at the heart of our control system is like our secret nuclear power reactor inside our brains. It is the “nerve center” of our being. It does *not* take existential threats kindly. Anything that appears to threaten our internal sense of identity and control gets raised to the red alert level. This is why you can't just “mansplain” something to someone else, and expect them to instantly see the error of their ways, and instantly become a new, better self. We have a lot of investment in that *old* self, and it does not look kindly on being deposed from its despotic rule over its own internal kingdom.

Although the self is a despot at heart, it is also remarkably sensitive to external, social forces, creating one of the most fundamental and puzzling paradoxes of the human condition: We care deeply about what other people think of us, and are actually remarkably malleable in adapting our behavior under the influence of others. There are many demonstrations of the power of the social force, from the evil of Nazi Germany and controversial attempts to recreate those forces in the lab, to the seemingly more benign and amusing phenomenon of hypnosis. Biologically and ecologically, our very survival is utterly dependent on our ability to work together socially, and social motivations are undoubtedly wired directly into the depths of our brains, providing these “hijack” pathways past the watchful eye of the self-model.

And therein lies the likely explanation for this paradox: these social forces can only act when delivered in ways that the self either does not recognize as threatening, or even endorses. The minute you are aware someone is trying to convince you of something, is the minute that it fails. But when a social virus is neatly packaged in a nice sugar coating, often in terms of reinforcing a sense

of belonging with an identified *in-group*, then it can easily slip past the guards. These kinds of in-group / out-group (tribalism) dynamics are the strongest of social forces and underlie all the greatest evils of humanity. And probably many of our greatest triumphs too.

Developmentally, the self emerges around age two, heralded by the onset of *tantrums*. Tantrums are the inevitable consequence of an emerging desire for control, coupled with an almost complete lack of *actual* control. This is really the defining battle of life, and it never really ends: the best you can hope for is some kind of truce as expressed in the Serenity Prayer of Reinhold Niebuhr: “God, grant me the serenity to accept the things I cannot change, Courage to change the things I can, And wisdom to know the difference.”

The Breakdown of Control

Unfortunately, achieving *serenity now* is very difficult. And all those challenges to the self can end up leading to a bout of depression, often coupled with anxiety or other unpleasant mental states. Although widely characterized in terms of *anhedonia* or the inability to experience pleasure, current research supports the idea that the core disorder of depression is really about *control*, or the perceived lack thereof. When your self model is sufficiently challenged, it basically gives up on a lot of goals, and unfortunately, achieving those goals is a primary source of pleasure and satisfaction in life. So, yes, anhedonia is a consequence of depression, but the core of it is more about the inability to motivate yourself to get out of bed and do all those now-meaningless things that you used to find meaningful.

Consistent with this central role for control, one of the most promising components of modern therapy for treating depression is *behavioral activation*, which is essentially an attempt to reboot your core self-motivation control system. Indeed a major study found behavioral activation to be the most important element among a group of therapies, and as effective as medication (Dimidjian et al, 2006). And when you recognize the central role for control in depression, it is then less surprising that medications are relatively ineffective: for the vast majority of people, the problem is *not* about some kind of low-level imbalance in their brain chemistry: it is about their core mental power plant running out of steam. And it just takes hard mental work, aided by effective therapeutic treatments, to reboot your own sense of mental self-control and efficacy.

For the smaller proportion of people who clearly do have a biologically-based mental disorder, it is still the case that the brain areas most centrally involved in self-control are the ones that are most likely to be affected. Schizophrenia and OCD for example involve the frontal cortex, basal ganglia, and dopamine systems of the brain, which are the main players in developing and sustaining our internal self control system. Thus, understanding how different parts of the brain function to support this critical self-model system is a major goal of current research in Psychology and Neuroscience, and this book is designed to get you started on a journey toward understanding this cutting-edge work.

Other Principles and Perspectives

There are many candidates for “the fourth C”, and different names could have been chosen to refer to the above “three C’s” (e.g., reduction, relativity, and... respect?), but being a slave to the simplifying force of *Compression*, it is useful to try to see as much as possible through the lens of these three principles. Furthermore, as briefly introduced above, these principles can be tied directly into the most fundamental properties of the nervous system, and thus provide a critical *bridging function* between Neuroscience and Psychology. Nevertheless, it is important to always remain aware of all the compressing taking place, and to acknowledge that this radical attempt at synthesis may strike many practicing scientists as overly simplistic or downright wrong-headed. However, my hope is that the benefits outweigh the costs overall, without attempting to overly minimize those costs.

Where do we go from here?

This question can be asked at two levels: the short-term question of where this book is headed, and the longer-term question of where our species is headed!? Although it may seem like our current cultural and political environment reflects an extreme magnification of many of the negative aspects of human mental function as described above, another perspective is that these truly are perennial battles and challenges that we have struggled with since the dawn of human history, and that they are borne of fundamental properties of the human brain that also have many positive aspects. Like everything it seems, double-edged swords abound. And the core premise, and promise, of science is that by understanding something deeply, we are better positioned to make the best of it. This contrasts with the idea that by somehow reifying “bad” features of the human brain, we are therefore justifying the bad ends they produce. Clearly that is not the aim here, and my personal optimism leads me to believe that this endeavor will be a net benefit in the end (or at worst, simply irrelevant).

With those big picture questions out of the way, we can turn to the plot for the rest of this adventure story through the human brain. Unlike a good mystery story, we’re going to ruin the whole thing right up front, in the hopes of achieving a better understanding and mental roadmap in the bargain.

Chapter 1 will provide a big-picture overview of the challenges and promise of achieving a scientific understanding of the brain and the mind (particularly the mind). The main challenge here is the *subjective* nature of the subject matter – however, we see that this subjectivity is actually primary and a fundamental challenge for all science, and a major challenge for people more generally!

Chapter 2 will cover the nuts and bolts of the brain, but always connected directly to the bigger picture via the three-C’s principles and their applications. We’ll see in detail how each neuron functions as such an amazing “information compactor”, compressing those 1000’s of signals into its single spiky output. We’ll then take an amazing “connected” voyage through the pathways of the

neocortex, seeing how the great chain of neurons locked in their long-lasting embraces create channels where information flows in different ways. We'll wrestle with the central question of whether brain areas are truly “specialized” for different functions or not, and whether there is any “there” there, as in, “where *is* that memory anyway?”

Chapter 3..

etc.

Chapter 1: Science and Subjectivity – The Fundamental Challenge of Psychology

Psychology is the science that attempts to understand the human mind. The human mind is the most fascinating and amazing “thing” in the known universe, and the idea that you can actually attempt to study it using the basic reductionistic approach of science may seem a bit of a stretch. And indeed it has been – but at this point in the development of the field, most practicing scientists are likely to feel rather confident that significant progress has been made, without fundamental, obvious limitations to how far we can go.

Despite all this progress and optimism, we will see in this chapter that there actually are fundamental boundaries to what science can penetrate, and these boundaries have shaped the field from its inception. Thus, understanding these limitations helps put the field of psychology and neuroscience into perspective in multiple ways, and in fact many of the limitations we discuss apply to science, and all human knowledge, more broadly.

The central issue we must confront head-on is the inescapable problem of *subjectivity*. By subjectivity we mean not just the fact that different people have different opinions or perspectives on things, though that is a big part of it. Instead, we need to step back a bit to look at the *really big picture* (i.e., Philosophy), starting with the fundamental problem of subjectivity as expressed by **Rene Descartes** (way back in 1637), in his famous statement: *Cogito Ergo Sum – I think therefore I am*.

There are two essential implications of this statement – we’ll explore the first one in depth before turning to the second. The first implication is that *subjective experience is primary*. If you put yourself into the mindset of a very skeptical, doubting philosopher, you might just about be able to get yourself to question everything, *except* this one, primary fact: you are sitting there (wherever you are), *thinking*. If you really push it, you might appreciate that you can’t really be sure that the world itself exists outside of your mind! This very challenging train of thought is well-captured in several modern movies, perhaps most notably in the *Matrix* series, where, in fact (in the movies at least), there turns out to be every reason to have such doubts. In philosophical circles, this line of thinking is known as *solipsism*, and lest you think that this is just an irrelevant and obscure way of thinking, one of the great innovators of our time, Elon Musk, is apparently convinced that we’re all living in a giant simulation.

This is the kind of all-encompassing subjectivity that we want to more fully understand and appreciate. What does this line of thinking mean for the study of psychology, or science more generally?

This is where we can usefully bring in Descartes’ second major implication from *Cogito Ergo Sum: dualism*. Dualism is the idea that there are two fundamentally different “substances” in the universe: the regular physical stuff of the everyday world, and this entirely separate, magical transcendent thing called *mind*, which lives apart from that other, regular stuff. The opposing view is

called *materialism*, where the mind is seen as just a product of the material world like everything else, and in particular a product of the physical processes taking place within the *brain*, as widely embraced in modern neuroscientific approaches to psychology.

You might be somewhat surprised to hear that many modern-day philosophers still embrace dualism, and one of the most outspoken advocates is David Chalmers, who argues that understanding the nature of subjective experience, or *qualia*, is the *hard problem* of consciousness and simply cannot be explained in objective, materialistic, scientific terms.

You might also be surprised to hear that, despite being one of those modern materialistic neuroscientists, I actually agree with Chalmers, and Descartes (in spirit at least, so to speak)! I think that there are two fundamentally different “somethings” in the universe, but, unlike Descartes and Chalmers, I don’t think the dividing line is between *mind* and *matter*, but rather, between *subjective* and *objective* perspectives.

Following Descartes (again), we can take subjective experience as primary – it is the only thing I am fully certain of. But it is also primary in another, essential way: it is uniquely, completely, definitionally, *mine*. It is literally impossible for *you* to experience *my* subjective experience, because, by definition, *my* subjective experience is exactly the sum-total of what it “feels like” to be me. If we somehow were to add *you* into my brain, my subjective experience would be irreparably altered. If you are somehow sharing in my subjective experience as it is happening, you would have to have direct access to every level of my brain, and not just “objective” access as you might get from a super-hi-tech future brain scanner, but *direct, internal, subjective* access, “from the inside out”.

In other words, you would have to literally be inside my brain. And you can’t be inside my brain because I’m already here. From the materialist perspective, we can identify my subjective experience as emerging directly from my brain – it is what it feels like to be my brain. If you truly appreciate this equivalence, then it should be readily apparent than there can be only one “mind” for every brain (we’ll look into the fascinating phenomenon of multiple personality disorder later, but it doesn’t change this fundamental conclusion – all those personalities are just as irrevocably trapped inside the one brain as you and I are, and in fact we all have something like multiple personalities too).

Another way of thinking about this is in terms of identical twins. Let’s imagine we have the most identical of identical twins ever to exist. Their brains are *completely identical* in every way possible. Would those twins have the same subjective experience? No. They might have a great deal in common, but, fundamentally, they would not, and could not, directly experience exactly what the other is experiencing. Why not?

It all boils down to *perspective*. Each physical thing in the universe has its own unique perspective, if we take this term to mean a particular spatial location, and a particular trajectory through space and time in the past (and going onward into the future), that is fundamentally *unique* to that thing. This

is why the twins cannot share their subjective experiences: they are two separate, distinct things, and, inevitably, they “see the world” from two different vantage points. The only way they could share experiences is if they could somehow superimpose themselves into exactly the same point in space, and do so over a sufficiently long time period to synchronize their history of experience, which plays such a critical role in our subjective life, in addition to the immediate sensations coming in from the outside world.

Anyway, the key point of all this is that *if* you allow that subjective experience can never be shared among different brains, *then* it follows that there is a fundamental divide between this inner subjective world, and the “regular” outside *objective* world. I believe this divide captures the essence of what Chalmers is talking about in terms of the irreducible nature of the qualia of consciousness – the impossibility of trying to explain in objective terms “what it feels like” to experience things in our subjective, inner world. Furthermore, it does so without introducing anything particularly magical or fundamentally at odds with materialism: subjective experience is not separate from the physical world in terms of some kind of magical “substance” that it is constituted from – it is just separate in terms of this notion of *perspective* – the unique point of view (literally, where they are standing / sitting / looking) that each subjective being has all to themselves.

Subjectivity in Psychology: A Brief History

Stepping back from this big philosophical abyss, what does it all mean for the attempt to study psychology as a science? The primary, obvious problem is that psychology is the study of *what it is like to be a human being*, and if this is fundamentally a subjective thing that can never be directly shared with any other human being, how can we possibly hope to arrive at some kind of objective, scientific understanding? Well, the first step is to follow Chalmers and attempt to *partition the problems* – we can carefully attempt to set aside the *hard problems* associated with the nature of subjective experience, and focus instead on the so-called *easy problems* that are left over. *If* there is enough interesting stuff left over in this space of easy problems, then it probably makes pragmatic sense to just see how far we can get in trying to understand that stuff, and then, once we seem to have exhausted that space, perhaps we could circle back and start reconsidering some of those hard problems.

This overall approach provides a reasonable narrative for the history of psychology as a scientific discipline. The person most widely credited with founding the science of psychology, **Wilhelm Wundt**, had the innovative idea in the late 1800’s that, after millenia of armchair speculation, you could actually apply the techniques of empirical science to understanding the human mind / brain. Wundt made many groundbreaking contributions, but his legacy, at least at the level of introductory psychology texts, is as a founder of the *introspectionist* school of psychology, which also includes **William James**, who also made major lasting contributions to the field. When the next major paradigm shift took

place in the early 1900's, it emerged as a strong reaction and rejection of this introspectionist approach, which was characterized as being overly concerned with all those hard problems of subjective experience. Introspectionists would try to systematize and characterize the contents of subjective experience, and the hard-nosed *behaviorists* who came next regarded these investigations as insufficiently objective, rigorous, and replicable. Instead, they emphasized purely objective, externally-observable *behavior* as the only valid data in psychology (hence the term behaviorism). The main figures in this era (e.g., **John B. Watson, B. F. Skinner, and Ivan Pavlov**) focused on how external, objective factors such as reward and punishment affected subsequent behavior through *conditioning*.

Thus, these first two epochs of scientific psychology embody exactly this tension between the subjective and objective worlds. The next paradigm shift took place in the 1950's and 60's with the *Cognitive Revolution*, riding the wave of digital computers, which made it fashionable to start talking about internal mental operations in terms of the *information processing model* of the mind – i.e., the mind as a computational device. Scientists leading this new field, such as **Herbert Simon** and **Alan Newell**, started thinking about how the mind could perform complex mental operations such as scientific proofs, chess, and other challenging tasks. People created running computer models of how these internal thought processes might work, which provided a compelling way to render that formerly “loosey-goosey” internal world in a much more rigorous, objectively-characterizable way.

However, as parallel work in the field of Neuroscience continued to advance, it gradually became clear that the brain really doesn't work anything like a standard digital computer. Instead, it is really a *massively parallel* computer with billions of computing elements (neurons) that combine the functions of computation and memory, which are otherwise separated in a standard digital computer. Psychologists **David Rumelhart** and **James McClelland** published a groundbreaking pair of books in the mid 1980's that popularized this new understanding of how information processing might work in the brain, and subsequent advances in the ability to take high-resolution pictures of the activity inside the human brain (*neuroimaging*) have led to the currently-dominant paradigm that integrates neuroscience and cognitive psychology (i.e., *cognitive neuroscience*) to come up with coherent understanding of how exactly the brain gives rise to the phenomena of the mind.

Fundamentals of Cognitive Neuroscience

This book is grounded squarely in this new paradigm of cognitive neuroscience, and attempts to provide a coherent set of core principles that connect directly from the basic processing carried out by individual neurons, all the way up to the highest levels of mental life. We are still largely avoiding significant consideration of the vast inner world of subjective life, but there is a robust field studying the *neural correlates of consciousness* (NCC) that we will discuss in depth in Chapter X. Slowly but surely, we are building bridges between the objectively-identifiable

properties of the human brain, and the subjective experiences that tend to co-occur with particular such brain states. Thus, we are developing a richer objective understanding about the kinds of neural mechanisms that give rise to our subjective mental life. But even with all of these advances, I don't think we could ever explain to a non-human-brain lifeform what it feels like subjectively to be a human brain. Thus, the subjective world remains our own private dominion, and probably literature, art, and movies provide the richest vehicles for sharing those experiences across the inevitable subjective gap between us all.

Subjectivity and Science: Working with the Method

The challenges imposed by the primacy of subjectivity have far-reaching implications beyond the field of psychology. First, given that some people can't even agree that there *is* an objective, external world outside the mind, how can we possibly even begin to start talking about *objective knowledge* and *facts*? This appreciation for the primary nature of subjective experience forces us to recognize that objective knowledge itself is entirely dependent on the subjective motivation of individuals to entertain a strong enough belief in this notion of objective reality, to put up with all the effort it takes to make any progress in understanding and advancing objective knowledge.

Those individuals are called "scientists", and they follow a particular method, the **scientific method**, which has the following basic steps:

1. Come up with a general question or problem, e.g., based on an informal **observation** about something of interest (e.g., Newton observes the apple falling on his head, which gets him thinking..)
2. Form a specific **hypothesis** about how that something might work, which makes testable **predictions** (e.g., there is an invisible force called *gravity* that causes all objects to experience the same acceleration, making the testable prediction that a feather and a hammer should fall at the same rate *in a perfect vaccuum* so as to eliminate the "confound" of friction).
3. **Collect data** that could actually test the predictions of the hypothesis, in comparison to other possible hypotheses (e.g., measure how fast things fall, ideally in a vaccuum if you happen to have one of those lying around). It is essential that the data be collected using a well-specified procedure that could be **replicated** by other scientists.
4. **Analyze the data** to determine whether any effects observed are strong enough to be clearly distinguishable from random chance and noise.
5. **Draw conclusions** – how compelling are the data, what holes are there in the data that would allow other hypotheses to explain the observed effects, etc?
6. Iterate! Plug the holes, think of other alternative explanations, test those, etc.

These steps can incrementally pull us out of our individual subjective fortresses through the critical lever of **consistency**. If you articulate a clear

sequence of steps to perform an experiment, and tell me exactly what you observe as results, and I do the same thing to the best of my ability, and get *consistent* results, then it seems like there might be something *real* and *objective* going on, or at least the world isn't completely random. As more and more people do the same thing, and continue to get consistent results, the odds that each one of us is just being individually tricked by some kind of subjective illusion would seem to go down.

As this scientific process continues, ever broader networks of interconnected hypotheses and associated empirical data accumulate, and if all of these remain somehow consistent with each other, it really starts to seem like there might be some kind of *laws* governing the behavior of the outside world. Furthermore, all this scientific knowledge makes its way into technology, which depends on those same laws, further bolstering the network of consistency. Fast forward to the modern world, and we now have the *standard model* of physics that provides a single consistent framework for understanding virtually all physical phenomena that have been subject to experiment, and drives incredible technology that would have been considered pure magic in times past.

Despite all this amazing progress made through the iterative application of the scientific method, you still have people like Elon Musk, one of the great *users* of physical laws, nevertheless concluding that it is all a giant simulation. And still plenty of people who believe that the Earth is flat, etc. And there is *nothing* you can do to convince these people otherwise. Such is the ultimate primacy of our subjective perspective on the world: the *only* porthole we have onto that supposed objective reality out there is through our very own, individual, subjective lenses. Because our subjective worlds are fundamentally uniquely our own, this also means that nobody can force anyone to believe anything that they aren't otherwise prepared to believe. Objective reality really is a second-class citizen, and is entirely dependent on the patronage of the ruling, sovereign subjectivity, just as scientists are still to this day dependent on the hard work and wealth of others to have the luxury of time and resources to create this huge network of consistent hypotheses and data.

Even within the scope of the scientific method, subjectivity abounds. Where, exactly are these hypotheses, or conclusions, supposed to come from? How many scientists looking at the exact same empirical data draw the same conclusions? You'd be surprised how subjective and inconsistent cutting-edge science really is. History is full of examples where a visionary pioneer was ridiculed by their colleagues, until enough evidence accumulated, and enough old people in power died, to allow the new ideas to flourish. The widely-accepted description of how science actually works, developed by Thomas Khun in 1962, emphasizes this sociological, psychological reality of science, with one major consequence being the strong suppression of ideas that are inconsistent with the current paradigm.

We can understand this phenomenon in terms of the three C's principles. Compression says that people crave simplicity, and the current paradigm embodies that: it is something that a large number of people know and agree about. Having that overturned requires confronting a high level of uncertainty

and complexity. Control is paramount here: that challenge to a widely-believed paradigm is experienced as a direct, personal challenge to your entire mental fortress – psychologically, it is really the same as challenging someone’s belief in a particular religion. Furthermore, the uncertainty directly undermines the feeling of control as well. And control interacts with contrast – the “paradigm believers” constitute a social in-group, and anyone challenging the paradigm is immediately a strongly-contrasting out-group member, and all the deep tribal motivations are aroused in this case, causing the challenger to be treated like a real outcast and pariah.

In other words, science is just people being people. However, despite all our limitations and inevitable subjectivity, there is some indication that following some approximation of the scientific method really does seem to work, at least over the longer arc of history.

Before we get more into the nuts and bolts of actual experiments and statistical analysis techniques in psychology and neuroscience, there is one further perspective on the problem of subjectivity in science that bears mentioning. This comes from Robert Pirsig, who wrote the famous book, *Zen and the Art of Motorcycle Maintenance*, which is actually more about philosophy of science and personal autobiography, rather than Zen per se. Pirsig literally went insane (as in, institutionalized, electroconvulsive shock therapy, etc) in the course of struggling with the question of where hypotheses come from – he realized that there was no rational explanation for how to come up with a good hypothesis, and it seems like there could easily be an infinite number of plausible hypotheses, so this throws a massive monkey wrench into the entire rational foundation of science.

Thus, subjectivity, creativity, and individual genius truly lie at the heart of science – most scientists are reasonably capable of evaluating hypotheses in terms of their consistency with data and with the larger network of other validated hypotheses, but relatively few scientists are responsible for coming up with the major hypotheses in the first place. Oh, and by the way, Pirsig suffered from Schizophrenia so that probably had more to do with his mental breakdown than the problem with hypotheses, but anyway it makes for a good story.

Research Methods in Psychology and Neuroscience

After all that philosophy, you might find a bit of concrete research methods a refreshing change! In this section, we’ll discuss the specific types of data that psychologists and neuroscientists tend to collect, and what kinds of analyses are typically done with that data. This is the kind of thing that almost everyone agrees about, and we will cover it very succinctly because it all sounds perfectly logical, but actually applying it requires a good deal of practice and experience, which is beyond the scope of this book, and likely the course you’re currently taking.

In psychology, there are three major ways in which data is collected, each

with complementary trade-offs:

- **Descriptive Methods** – these tend to be the least *invasive* techniques, involving various ways of capturing what is actually happening in human behavior, such as observation, case studies, and surveys. A modern version employs cell phones with apps that ping people at random times during the day and ask them what they’re doing, or thinking about, etc. The disadvantage of these techniques is in their relative inability to inform you about *why* people might be behaving the way they are – the other two techniques improve on that aspect of things, but, particularly with the experimental method, tend to require more artificial, less naturalistic kinds of experiments.
- **Correlational Studies** involve measuring multiple different **variables** (something that can be measured which varies across people, such as weight, IQ, vocabulary, diet, etc) and determining the extent to which these variables **correlate** or vary systematically in relationship to each other. For example, people’s weight and height tend to be positively correlated, because as one goes up, the other does too. Critically, as with most real-world data, this is not a **perfect** correlation – there are many exceptions in either direction – but overall, on average, there is a relationship. The single most important limitation of correlational studies, is that the presence of a **correlation does not imply causation**. Typically, causation does imply correlation of some sort, but this relationship is not symmetric! Unfortunately, the human brain relies on correlation as a kind of “quick and dirty” shortcut for finding causal relationships in the world, and we find it remarkably difficult to recognize that the two are not equivalent. For example, most studies on the effects of diet on health are correlational, and yet the media and even scientific papers regularly interpret these as showing a causal link. “Drink more coffee because you’ll live longer!” Well, what if in fact the observed correlation between coffee and longevity is due to the fact that more wealthy people drink more coffee, and it is really the wealth and all its associated benefits that is driving the longevity. Coffee is just “along for the ride”. This is the **third variable problem** (in this case, the third variable is wealth), and it is the bane of correlational studies, because *there is always a third variable* (and a fourth, and a fifth, etc). And it is typically very difficult to rule out the possibility that everything is being caused by one of these unmeasured “third variables”.
- **Experimental Studies** are the only way to truly establish a causal relationship, and even then it is still a major challenge to really accomplish this feat. The key trick is to use randomness and careful designs to attempt to systematically eliminate all possible “third variables”. A huge source of third variables is each individual person participating in the study. Like all the bacteria on your skin, you are crawling with third variables. Your genes, your upbringing, your neighborhood, your schools, your friends, your... everything, is a teaming cesspool of third variables! The key trick in an experimental study is to use the cleansing power of randomness to

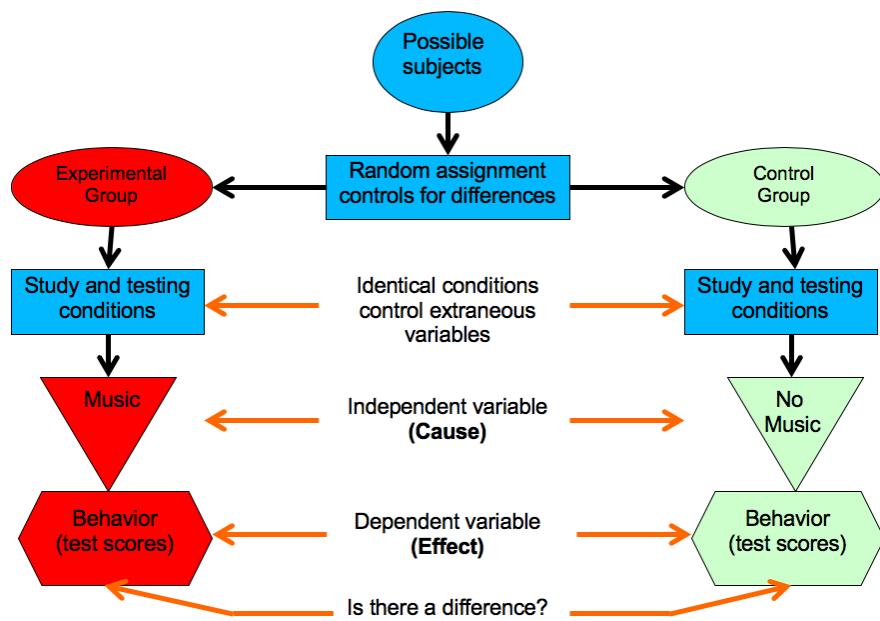


Figure 2: Fig 1-1: Logic of an experimental study, using random assignment to eliminate third variables from the study participants. It is also essential to minimize all other differences between the experimental and control conditions (i.e., *confounds*, or additional “third variables”), to more precisely identify the single *independent variable* (i.e., the *causal variable*) as truly being responsible for the differences measured in the *dependent variable*

wash away all those third variables, by **randomly assigning people to different conditions**. No third variable can withstand the incredible power of such random assignment – if we find a systematic difference between two completely random samples of the population, it cannot be due to their pre-existing conditions! However, random assignment is also the achilles heel of experimental studies, because it is often impossible to use random assignment for many questions of interest. Can you really look at the effects of parenting style on subsequent emotional development, by randomly assigning kids to parents!? Same goes with any long-term study on things like diet and lifestyle – you can sometimes sorta force people to eat some particular diet over a period of a few months or so, but that just isn't going to work for the decades it likely takes for most diet effects to really impact overall health outcomes. There are also other important ways of eliminating further possible third variables (typically called **confounds** in this context) from experiments, but random assignment is the most important (see Figure 2-1 for a diagram of the overall logic).

Thus, each of these different techniques is most appropriate for different kinds of questions, given the different tradeoffs. The key thing as a student and a citizen is to understand the limitations of any given study, so you can make an informed decision about what it really means. And don't expect the media to do this for you. Seriously, look at *any* correlational study on health / diet / etc and see how clearly the story, or the original article, discusses the limitations on any kind of causal implications from the study.

Neuroscience methods

Methods in neuroscience (and cognitive neuroscience) tend to be either correlational or experimental. The vast majority of **neuroimaging** studies are purely correlational, measuring the neural correlates of various different tasks or other manipulations performed while participants are in the brain scanner. By now, the neural correlates of just about every possible human activity (yes, including sex) have been measured in a scanner. But because of the correlational nature of these results, it is difficult to know whether the recorded brain activity is just *epiphenomenal* (i.e., just along for the ride), or whether it is really causal and somehow *responsible* for the behavior in question.

To attempt to address this causality question, scientists have used various forms of electrical and magnetic stimulation, which can disrupt or enhance neural firing in a relatively localized region of the brain. For example, **transcranial magnetic stimulation (TMS)** applied over the primary motor cortex can cause your muscles to flinch. However, just as with other experimental studies, the resulting brain states after TMS are not very “naturalistic”, and it becomes difficult to interpret whether any changes in observed behavior are due to the disruption of the “normal” functioning of that brain area, or whether they just reflect the weird stuff that happens when you tweak that brain area in a completely unnatural way.

In animal neuroscience, much more precise causal inferences can be made by employing much more “invasive” techniques, such as directly cutting out different parts of the brain, or using modern **optogenetic** techniques to instantly and reversibly activate or deactivate a given population of neurons. These optogenetic techniques allow very specific populations of neurons to be targeted, and have produced a powerful new wave of causal empirical data, showing that very precise manipulations to very specific neural populations can sometimes have impressive overall effects. However, often even these results are over-interpreted and one must look very carefully for confounds in the resulting activity of other neural populations. Virtually every neuron in the brain is within a few synapses of every other neuron (i.e., the “6 degrees of separation” (from Kevin Bacon) phenomenon), so it remains very difficult to isolate what each specific subset of neurons is uniquely contributing. Indeed, as we’ll see in the next chapter, the very premise of isolating specific functions may be entirely misguided.

Finally, animal neuroscience also affords much higher-resolution neuroimaging techniques which can resolve the activity of individual neurons, while also recording many such neurons at the same time. Such techniques provide the most powerful descriptive methods for characterizing what neurons actually do, and historically have been some of the most important data for fueling our theorizing and understanding of how the brain works.

Thus, truly each different type of technique plays a critical role in the overall arsenal of science.

Statistics

Finally, it is useful to be aware of the most widely used statistical techniques in psychology and neuroscience. Here is a brief overview:

- **Descriptive Statistics** – like descriptive methods, descriptive statistics are used to describe data, and differ from **inferential** statistics which are used to *infer* causality or correlation, as described below. The primary descriptive statistics are probably familiar to you: *mean, median, mode, range* and *standard deviation*. For a *normal* (bell-shaped, *gaussian*) distribution, the mean, median, and mode are all the same, and they tell you where the *middle* of the distribution is (i.e., the “average” person, etc). It is only when the distribution is *skewed* that they differ, with the mode and median being less “pulled” by the long-tailed side of the distribution. You may have heard of income being reported in terms of medians – this is because income is a skewed distribution, with progressively fewer people making a *lot* more money than the mass of the “middle class” and below. The median and the mode more accurately capture this “middle class” salary because they don’t get pulled upwards as much by all the rich people.
- **Correlation Coefficient and Scatterplots** – these are the primary tools for correlational studies. The correlation coefficient is a number, typically labeled *r*, which goes between -1 and 1, where -1 represents a

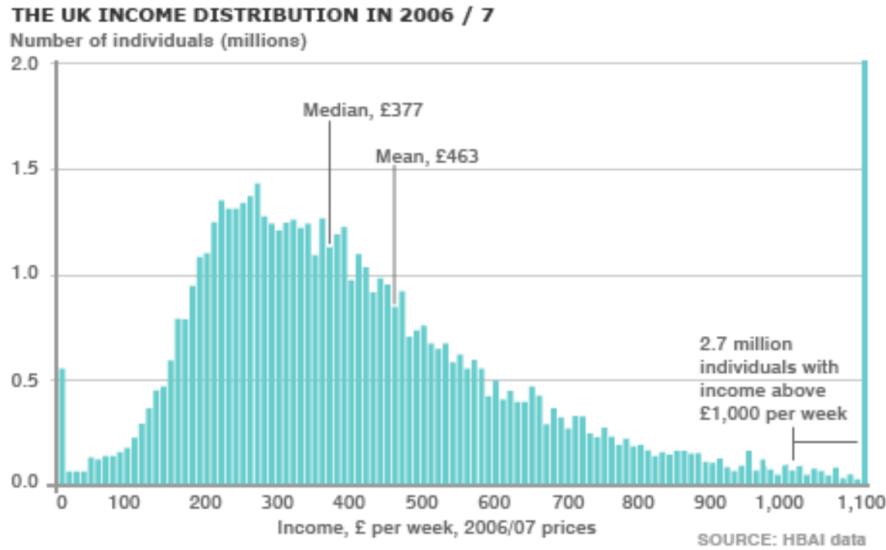


Figure 3: Fig 1-2: Mean, Median, and Mode tell different stories when the distribution is skewed (in this case, it is *right-skewed* – the skewer is the long tail to the right). The mean is pulled up by the tail much more than the median or mode, which do a better job of capturing the “middle class” income.

perfect negative correlation, 0 is the complete absence of a correlation, and 1 is a perfect positive correlation. Importantly, both a strong negative and a strong positive correlation are equally important statistically, and indeed you can almost always just flip one of your variables around and turn one into the other (e.g., height vs. weight is positive, but “shortness” vs. weight is negative). A scatterplot simply plots the value along each variable (one on the X or horizontal axis, and the other on the Y or vertical axis), with each dot representing a different person (or whatever else is being measured). Thus, you can usually directly see the strength of the correlation in the shape of the “cloud” of such points (todo: include standard figs). One critical “pro tip” for looking at such scatter plots is finding “outlier” points that might be carrying a huge amount of weight. Just as a person sitting further out on a see-saw has more impact than one sitting further in, data points that are far away from the center of the cloud carry much stronger weight, and if they happen to lie along one of the positive or negative diagonals, they can produce a strong apparent correlation, even when all the rest of the points in the middle are clearly just milling about and going nowhere in relation to each other.

- **t-test, F-test (ANOVA) and the GLM** The “Student’s” t-test is the most basic of the *inferential* statistics used in experimental studies. It is *not* so-named because it is only for use by students, but rather it was the

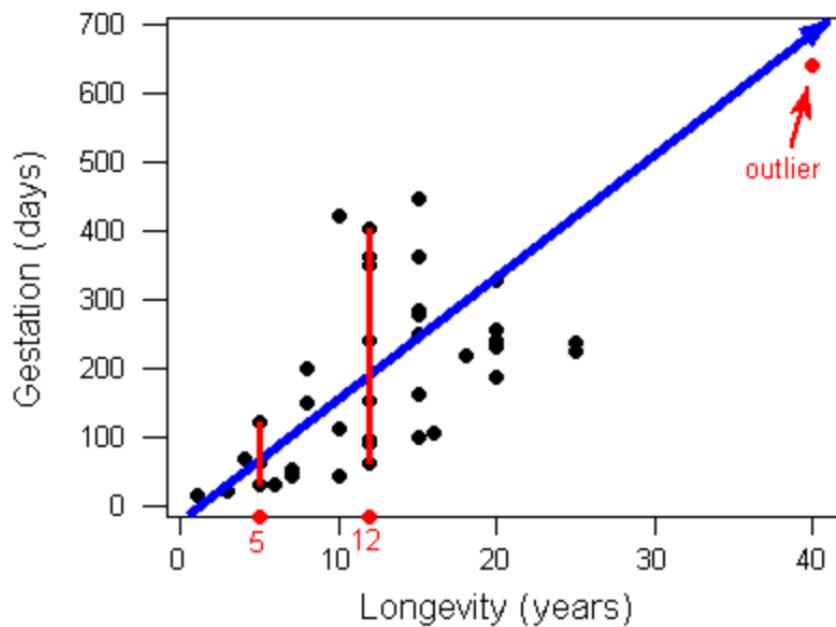


Figure 4: Fig 1-3: Scatterplot showing the positive correlation between length of gestation in the womb and overall lifespan, for different species of animals. The Elephant in the figure is the outlier, carrying undue amount of weight on the overall correlation coefficient. In this case, it is actually consistent with the rest of the data, but sometimes it is not, and yet the correlation still looks positive according to the r value. Thus, it is *essential* to *always* plot your raw data and ensure that the summary statistics are reflective of real aggregate effects!

pen-name of the guy who invented it (William Gosset), to improve the quality of beer brewed by Guiness brewery in Ireland, no less! Too bad it isn't called the "Stout" t-test. Anyway, it basically tells you if the difference between your experimental group and your control group is big enough to *not* be due to random chance. Thus, in applying this test, we "reject the null hypothesis" that our data is just random noise, but, critically, we're not actually *proving* that our favored hypothesis is correct. We're just saying it is relatively unlikely to be pure noise. There are more "advanced" versions of this test, specifically the F-test used in the ANOVA (analysis of variance) procedure, and the full *generalized linear model (GLM)*, which can tell you about the importance of multiple different factors and their potential interactions. You may have heard about the *replicability crisis* in various fields of science, including psychology, where many results that were thought to be "true" have "failed to replicate" – meaning that the original paper(s) reported a *significant* t-test result, and the subsequent ones did not (they instead found results consistent with pure noise). This is actually to be expected about 5% of the time, given the standard for publication is set at this 5% level. However, when you take into account how science is *actually* done, there are major systematic biases that enter into the process, which are not taken into account by these statistical tests, such that the actual effective probability of publishing garbage is closer to 50%! There are now important changes afoot to combat the worst of these biases, and help ensure that this garbage probability goes back down to closer to 5%. But 5% itself is still a rather large number – in physics the standard is one in 3.5 million! And, amazingly, results that end up going into the "garbage" pile appear significant at levels below this standard, so randomness can sometimes be a challenging foe.

Chapter 2: Neuroscience

From a materialist, neuroscientific perspective, *everything* that happens in your mind is due to underlying physical processes taking place in your brain. As we discussed in the last chapter, this does *not* mean that we can *reduce* your mind to the brain, but it does mean that there is a really huge mystery here: how is it even remotely possible for a physical system to produce the amazing subjective delights (and terrors, and everything in between) that we all experience?

We start with a time-honored scientific approach: reduce the problem to the simplest possible system that exhibits the relevant behavior, and see if that makes it easier to understand. Consider the two gears as shown in Figure 2.1. As elaborated in the figure caption, there is something kind of “magical” that emerges out of the interaction between the two gears, which cannot be reduced directly to either gear separately. These **emergent** properties depend critically on the relationship and interaction between the two different parts – their relative sizes, rotational speeds, etc. If the larger gear interacted with a different, even larger gear, the overall system of interacting gears would exhibit very different emergent properties. Thus, you really can’t isolate these emergent properties to either gear in isolation. Furthermore, the actual material that the gears are made of is largely irrelevant, as long as it is reasonably solid. Thus, there truly is some kind of seemingly mysterious new “substance” being created out of this interaction, which can “transcend” its material basis. And yet, it nevertheless depends entirely and directly on having an actual material basis – the *picture* of those gears doesn’t work at all like two actual gears!

This simple two-interacting-gears scenario captures the strange relationship between mind and brain, where the mind depends entirely on the brain, and yet it fully transcends it. As we’ll see in a moment, the brain has billions of tiny, interacting parts (*neurons*), which, like the gears, interact in ways that produce emergent properties transcending their material substance. Moreover, there are so many neurons in the brain, and each one interacts with so many *other* neurons (receiving roughly 10,000 inputs and sending a similar number of outputs), that there is a vastly greater degree of emergent interactions taking place in the brain compared to our simple gear example. Thus, although it is essentially impossible for us to wrap our own minds around it, it should be possible to at least imagine in a vague way how something as fantastic and complex as the mind could indeed emerge out of all those billions and billions of interactions taking place every nanosecond, right inside your very own brain.

To try out another metaphor, you can also think about the brain as a massive LEGO set, with parts that *learn* to interconnect with each other in myriad ways. As you might have experienced in your youth, the number of different ways even a small pile of LEGO’s can be combined to make different things quickly exhausts the imagination. This *combinatorial explosion* of possibilities is an essential feature of the brain – our neurons can be interconnected in so unimaginably many different ways, that the possibilities are effectively infinite. Due to the explosive nature of combinations, even small numbers of elements can be combined in

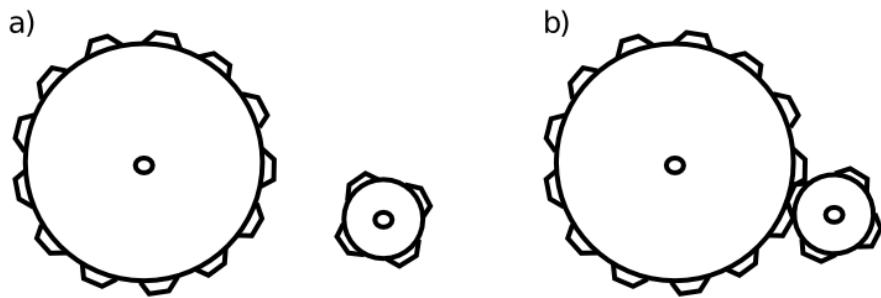


Figure 5: Fig 2-1: The principle of *emergence*, simply illustrated. The gears on the left do not interact, and nothing interesting happens. However, on the right, the interaction between the gears produces interesting, useful phenomena that *cannot* be reduced to the individual gears *separately*. For example, the little gear will spin faster, but the larger one will have higher torque at its axel – these properties would be entirely different if either gear interacted with different sized gear. Furthermore, the material that the gear is made from really doesn't matter very much – the same basic behavior would be produced by plastic, metal, wood, etc. Thus, even in this simple case, there is something just slightly magical and irreducible going on – when two gears get together, something emerges that is more than the sum of the parts, and exists in a way independent of the parts, even while being entirely dependent on actually *having* those parts to make it happen. This seems like a good analogy for the relationship between the mind and the brain.

more different ways than there are atoms in the universe. To see this for yourself, type in 69! (factorial) on your calculator (or just google it), and you'll get a number that is 1.7... with 98 zeros! This factorial function lies at the heart of combinatorial explosion, and gives a rough sense of the number of different combinations of 69 parts. You can't even begin to conceive of (or even calculate) the value of 100,000,000,000! (i.e., the factorial of the 100 billion neurons in your brain).

Simple Neurons Make Complex Work

One of the secrets of LEGO is that all the different parts interconnect using a single, simple principle, so that you really can make all those different combinations work. The same is true of the brain: each neuron operates according to surprisingly simple, easily-understood principles, and the power emerges through all the interactions / combinations of these simple parts. For full disclosure, not everyone agrees with this perspective, and many scientists have spent a long time exploring more complex kinds of languages that neurons might speak, but that is a topic for a more advanced textbook. Through the use of computer models of neurons, we can at least say with confidence that the simple ideas presented here can account for a surprisingly large amount of what we know about brain function.

The easiest way to understand what neurons are doing is in terms of **detection** (Figure 2-2). A neuron acts much like a smoke detector, constantly sampling its local environment (i.e., its inputs from other neurons), and looking for some set of incoming signals that indicate that something *important* might be going on (e.g., a fire in the case of the smoke detector). When it detects whatever it is looking for, it sends a signal out to other neurons, alerting them to the news, so they can incorporate this as one of many other pieces of information that they are sampling in their own detection process. And so on, and so on... Each of the important biological parts of the neuron take on a clear functional role within this overall detector model:

- **Synapses:** are the tiny gaps between neurons, where the output signals from one neuron cross over and become the input signals to the next. Most synapses are *chemical*, involving the release of a *neurotransmitter* from the *presynaptic* axons, which then bind to *receptors* on the *postsynaptic* dendrites. These receptors twist open as a result of neurotransmitter binding, and allow *ions* (i.e., electrical charge) to flow into the dendrites, through the resulting open channels. The most common neurotransmitter in the neocortex is *glutamate*, and it opens up *AMPA* receptors, that allow sodium (Na^+) ions to flow into the dendrites, thus creating a *positive* electrical potential in the receiving neuron. All this complex-sounding machinery and terminology is actually very simple: Neurons like to excite other neurons by sending them exciting signals! The basic machinery is chemical and electrical, but the bottom line is just: how strongly do the input signals to a given neuron excite it? This is determined by the detailed

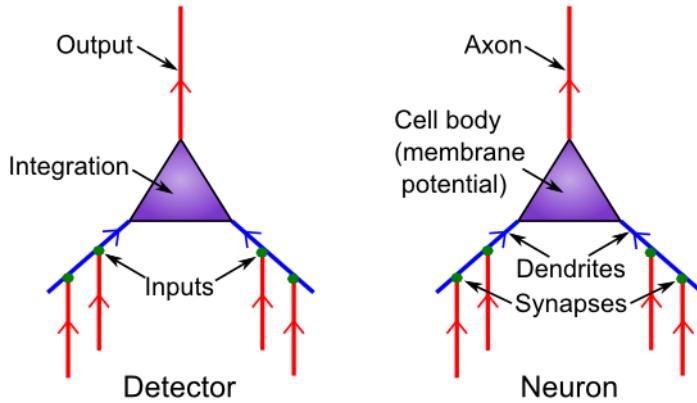


Figure 6: Fig 2-2: The neuron as a detector. Inputs come in via *synapses* connecting the *axons* of other neurons to the *dendrites* of a given neuron. This neuron *integrates* these inputs , resulting in an overall *electrical potential* (called the *membrane potential*, because it is the electrical difference between the inside and outside of the neuron's cell membrane), in the cell body. At the start of the axon (the *axon hillock*), a critical *go / nogo* “decision” is made – if the membrane potential is sufficiently elevated, then the neuron triggers an *action potential* (aka a “spike”), which races up the axon and delivers its signal to the many thousands of other neurons that are “listening” to this signal, via their own synaptic connections. Thus, the essence of neural function is *communciation* – neurons are highly social little things, and our brain is really a huge social network of chattering naybobs.

function of each of the roughly 10,000 synapses coming into a given neuron.

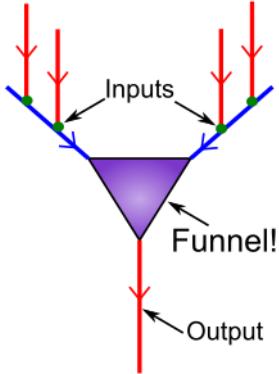


Figure 7: Fig 2-3: The neuron as a *funnel*, *compressing* its 10,000-odd inputs down into a *single* output signal, conveyed through its axonal output. This is the genesis of the *Compression* principle of brain function.

- **Dendrites:** provide a broad tree-like (dendrite literally means tree-like) “arbor” for all the synaptic inputs into a neuron, and they funnel the resulting electrical charges up into the cell body. This funnel-like property, illustrated in Figure 2-3, is the origin of the **compression** principle of brain function, one of our three C’s. As we noted in the introduction, each neuron is compressing its 10,000 different inputs into a single output signal, producing a roughly 10,000-to-1 compression factor. As an aside, one of the big debates in neuroscience is the extent to which these dendrites perform various kinds of more complex “processing” of their synaptic inputs, or simply convey the overall signal. There is evidence on both sides, and, as usual, the truth is likely somewhere in between.
- **Cell Body:** The neuron is a cell, and, despite its long tendrils, it has a cell body like other more compact kinds of cells, where the nucleus and other cellular machinery hangs out. It is here that all the dendritic signals converge, to produce the final compressed electrical potential that somehow summarizes everything coming into the cell at that moment. This electrical potential is called the *membrane potential* because the electrical signal is measured as a difference in electrical potential across the cell’s membrane (that fatty lipid bilayer that you might recall reading about in high school science). If this membrane potential is sufficiently excited, then special channels at the start of the axon (the *axon hillock*) will get *extra* excited and essentially flip a switch, causing the initiation of the *action potential* or *spike*. The details of this process were worked out by Hodgkin and Huxley in the 1950’s, and have stood the test of time, forming the basis of modern detailed mathematical models of neuron firing.
- **Axon:** The spike propagates down the axon, effectively broadcasting this

one signal out to the roughly 10,000 other neurons that it sends input to, continuing the great chain of communication among neurons. Axons can have varying amounts of *myelin*, provided by helpful *glia* cells called *oligodendrocytes* (no you won't be tested on those!), which serve to insulate the electrical "wire" that is the axon. Myelinated axons convey information more quickly, and *multiple sclerosis* is one of various disorders that involves the degeneration of this myelin, resulting in slowed signal conduction. There are many other forms of glia cells, but all of them are generally thought to play various supporting roles in the overall function of the brain, whereas the neurons are the "stars" of the show. These supporting roles are essential for keeping the brain functioning, and may affect various processes such as learning, but we'll nevertheless generally ignore them in this introductory treatment.

To summarize, each neuron is receiving a huge amount of input through its roughly 10,000 synapses, and it then compresses this all down into a single discrete spiking signal that it then broadcasts back out to the roughly 10,000 other neurons listening to its little story. Only when a neuron detects something "interesting" does it get excited enough to send this spiky signal out. This *thresholding* is just as important in neurons as it is in people: it can rapidly get tiresome to listen to someone with a *low threshold* who is always blabbering on about the most uninteresting things. The advent of Facebook and other forms of social media has only exacerbated this problem (which indeed did exist prior to such modern tech, as I can personally attest).

We'll explore more about what kinds of interesting conversations neurons might be having in a bit, but first we'll examine how this electrical magic operates within the neuron in more detail. These details are typically not presented at this introductory level, but a really simple analogy helps make it accessible, and this machinery ends up producing the **contrast** effects that are so central to our overall framework, so we're motivated to take this brief detour.

The Tug-of-War in Your Brain

There are actually two major classes of synaptic inputs converging on each neuron: the excitatory ones described above (via the neurotransmitter glutamate opening AMPA receptors), and separate *inhibitory* synaptic inputs that are driven by a neurotransmitter called *GABA* which activates... *GABA* receptors! These inhibitory inputs come from an entirely separate set of specialized neurons known as *inhibitory interneurons*, which are somewhere between the principal, excitatory neurons and the glia in overall status within the pecking-order of the brain. These interneurons only act relatively locally, like glia, and they also play a largely *regulatory* role, regulating the overall level of electrical excitation surging through the brain. In contrast, the main excitatory neurons (which constitute roughly 85% of neurons in the neocortex) can broadcast their exciting messages over long distances to far-flung regions of the brain, and are regarded as the primary *information processing* neurons (i.e., they are primarily responsible

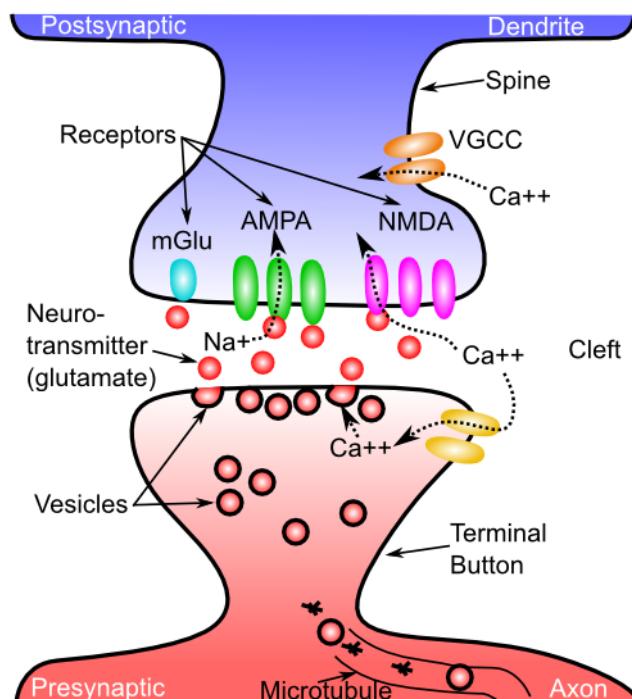


Figure 8: Fig 2-4: Details of the communication process across the synapse between a sending axon and a receiving dendrite. Neurotransmitter is released from the *terminal button* (pronounced in French-style by those in the know), and opens corresponding *receptors* on the dendrites, allowing electrically-charged *ions* to flow into the receiving neuron.

for all the chatter and compression going on).

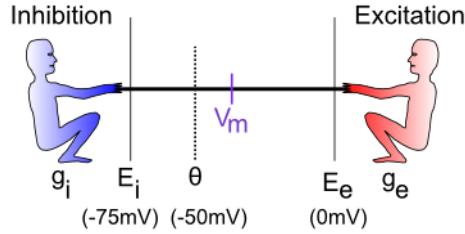


Figure 9: Fig 2-5: The Tug-of-War between excitation and inhibition, producing a beneficial balancing act, and the source of **contrast** coding in the brain. Inhibition pulls the membrane potential (written as V_m , where V =voltage and m =membrane), down toward the *resting potential* of roughly -75 mV, while excitation pulls up toward roughly 0 to +55 mV (depending on type of neuron, etc). Theta represents the *threshold* electrical potential, above which the neuron will fire a spike. The ability to do so depends only on the *relative* balance between excitation and inhibition, not the absolute levels.

Inside each neuron, excitation and inhibition are forever locked in a pitched battle, which can be pictured as a tug-of-war, with each side pulling with varying strength, but always in the same direction (Figure 2-5). The “pitch” on which this battle is taking place is the amount of electrical charge in the cell, i.e., the membrane potential. The excitatory end is always pulling this potential upwards, while the inhibitory side is pulling it back down, and the actual potential represents the balance between these two forces.

The essential point here is that **contrast**, or **relativity** emerges as the result of this tug-of-war battle. Specifically, it doesn’t matter how strong the two different sides on the tug-of-war are in absolute terms – all that matters is the *relative* strength of the two sides. Excitation could be relatively weak, but if inhibition is also weak, then the net balance between the two will be the same as if each was proportionally stronger.

Typically, the amount of inhibition is roughly proportional to the “average” amount of activity in the brain in any given area, so in effect, each neuron is effectively comparing how excited it is against this overall “average” level. Only those neurons that are getting *above average* level of excitation will actually get excited enough to fire spikes.

In real-world terms, this “average” inhibition is very much like the amount of money that your peers are making (or the amount of fun they appear to be having on their various social media accounts) – it forms the baseline or standard against which you measure yourself. Likewise, neurons are constantly comparing themselves against *their* peers, and all of the spiking going on in your brain is therefore always and inexorably *relative* to these peer-standards. For example, when you step outside into the bright sunlight, all the visual neurons suddenly

get a huge wave of excitation relative to the dim indoor light from before. But you avoid suffering an epileptic seizure from all that excitement because those inhibitory interneurons are also getting this wave of excitation, causing them to send a proportional amount of damping inhibition on the party, keeping everyone in balance and on a more level keel. Yes, inhibition is the wet rag of the brain, but without it, you really would be suffering from seizures all the time.

In fact, this balancing act between excitation and inhibition is so important for overall brain function, that our brains are perched on a kind of “knife edge”, and the relatively high incidence of epilepsy in the population is likely a result of the fact that it is really hard to get this balance exactly right. And too much inhibition has very bad consequences as well (indeed, it literally “depresses” your brain and makes it difficult for you to do anything). Furthermore, the main treatments for epilepsy involve activating the GABA inhibitory system more strongly, thus altering this fundamental tug-of-war balance.

In summary, two out of the three of the core principles of this textbook, **compression** and **contrast**, emerge directly out of the basic function of neurons. As we discussed in the Introduction, we can trace the implications of these core neural properties all the way through the full scope of Psychology and behavior. The next chapter will provide particularly compelling demonstrations of how compression and contrast play out in our perceptual lives – the story of perception is fundamentally the story of compression and contrast.

By thinking in these terms, we have managed to dramatically simplify our understanding of the brain, creating an almost transparent, level-spanning way of going from single neurons on up. However, none of this contradicts the emergence and complexity discussed at the outset. Instead, these principles just capture the overall general tendencies and propensities of the brain, but within that broader scope, there is a wild, complex, bubbling jungle of intertwined conversations and chatter constantly unfolding within your brain, thinking all manner of complex and ineffable thoughts. Next, we’ll move up from the level of individual neurons and start to think about how all these principles might play out in terms of how different brain areas are organized to facilitate effective overall behavior and cognition.

Large-Scale Brain Organization (“Gross” Anatomy)

“All politics is local” is as true for the brain as it is for people, perhaps more so: neurons can’t pack-up and leave. They are like the 85% of Pittsburghers who live their entire lives in the same little neighborhoods. This creates the possibility of *specialization of neural function* – different neighborhoods can develop their own special “personalities” and focus on detecting particular kinds of signals. On the other hand, the excitatory neurons in the brain also send out long-range connections. Network theorists characterize these as “small world” patterns of connectivity, such that, in the end, every neuron is only a few synapses away from every other neuron. This then limits how much “neighborhood funkiness” can develop.

Historically, there was a classic debate between Karl Lashley and ?? about this question of specialization of function, with Lashley showing that he could lesion large amounts of just about anywhere in the rat brain, and produce similar-looking behavioral effects. However, it turns out that he was using very simple behavioral tests, and when you push those poor brain damaged rats harder, they actually do fail in different ways with different types of brain damage. But not perfectly, systematically so. Again, the truth is always somewhere in the middle of such debates.

todo: detailed brain areas, follow the connections, etc.

Acknowledgments

Thanks to the current beta-testers for reading!

Glossary

About the Authors

Randall C. O'Reilly is Professor of Psychology and Neuroscience at the University of Colorado Boulder.

References