CHAPTER THE FIRST

DEALING WITH THEORETICAL DIVERSITY[1]

ÆTIOLOGIES OF DEVELOPMENTAL DYSLEXIA AS A CASE IN POINT

> "Hypotheses are nets; only he who casts them will catch. Was not the discovery of America the result of a hypothesis? Long, and above all, live the hypothesis – Only she will stay forever new, no matter how often she defeats herself. […] The skeptic, my friend, has done so little, and vulgar empiricism even less, for the extension of science. All the skeptic really does is spoil it for the hypothesisers, make them stand on shaky ground; a strange way of advancing science. At best, it is a very indirect achievement. The true hypothesiser is none other than the inventor, because before he actually makes his discovery, the discovered land already floats obscurely in front of his eyes. He carries this faint image into every observation and every experiment and at length –by means of free comparison of repeated contact and collision of his ideas with experiment– finally arrives at the idea, which stands in negative relation to positive experience. Both are then eternally connected and a new and celestial light surrounds the power that has come into the world."
>
> – Novalis (1798)

1.1 THE ELUSIVE ÆTIOLOGY OF DEVELOPMENTAL DYSLEXIA

When a child fails to acquire a proficient level of reading and spelling performance after about two years of regular literacy education in their native language and such possible causes of delay as below average intelligence, a specific sensory or motor deficit, or a general learning disability have been ruled out, the diagnosis is usually developmental dyslexia. Though many dyslexic readers will benefit from a timely intervention, the impairment will persist into adulthood with varying severity. Estimates of people afflicted with this specific learning disorder vary from country to country and ranges from 5-20% of the population (Blomert, 2005).

The definition of developmental dyslexia provided above is based on exclusion criteria. It describes the inability of otherwise typically developing children to acquire a proficient level of reading and spelling after regular literacy education (here typical just means average or normal with respect to development). Of course, there are more sophisticated diagnostic definitions that include a genetic component and mention a neurological basis for the impairment (cf. Fletcher & Lyon, 2008). There are also many more exclusion criteria based on results from elaborate test batteries and comorbidity with, for example, ADHD or Autism (e.g., Blomert, 2005). All these definitions remain however descriptive (or statistical) in nature and are essentially based on exclusion criteria. The practice of including genes or a brain region or function as an essential part of the definition appears to provide more precision or a deeper understanding of what causes developmental dyslexia, but that is just illusory precision. To date, there is no specific genetic profile (cf. Grigorenko, 2001), no specific brain structure or function (Caylak, 2009; Eckert, 2004, 2010) and no specific psychometric test profile (Blomert & Vaessen, 2009; Ramus & Szenkovits,

---

[1] This chapter may be cited as:

Hasselman, F. (2014): Dealing with theoretical diversity: Ætiologies of developmental dyslexia as a case in point. In *Beyond the boundary. An analysis of verisimilitude and causal ontology of scientific claims: Aetiologies of developmental dyslexia as a case in point*. fig**share**. http://dx.doi.org/10.6084/m9.figshare.1064427

The supplemental materials to this chapter may be cited as:

Hasselman, F., (2014). Beyond The Boundary - 1st Chapter: Supplemental Materials. Retrieved from Open Science Framework, https://osf.io/8y4sq

2008; Wimmer & Schurz, 2010) that can serve as the ultimate diagnostic tool. So what is it that is so specific about this specific learning disability, other than the observed problems with acquisition of proficient reading and spelling ability?

A quick survey of recent literature pertaining to the performance of dyslexic readers on a wide variety of tasks and experimental conditions seems to suggest that there is always some sample of dyslexic readers in some part of the world that can be found to deviate from average performance. There now exist a plethora of theoretical accounts of developmental dyslexia that explain such observed deviations from average performance as impairments in low-level sensorimotor processes or high-level cognitive processes, or both. The reported deficits span almost **all modalities of perception** (Beaton, Edwards, & Peggie, 2006; Breznitz, 2003; Goswami, Fosker, Huss, Mead, & Szűcs, 2010; Huss, Verney, Fosker, Mead, & Goswami, 2010; Skoyles, 2004; Talcott et al., 2003; Tallal, 2004), include **deficits in motor control** (McPhillips & Jordan-Black, 2007; Nicolson & Fawcett, 2006; Ramus, Pidgeon, & Frith, 2003; Savage, 2004), **balance** (Rochelle & Talcott, 2006; Stoodley, Fawcett, Nicolson, & Stein, 2005), **attention deficits** (M. Reynolds & Besner, 2006; S. E. Shaywitz & B. A. Shaywitz, 2008; Valdois, Bosse, & Tainturier, 2004), **impaired cognitive abilities** (Aleci, Piana, Piccoli, & Bertolini, 2010; Heim et al., 2008; Helland, 2007), **fluency of naming** (Araujo, Pacheco, Faisca, Petersson, & Reis, 2010; Vaessen, Gerretsen, & Blomert, 2009), **learning** (Menghini, Vicari, Mandolesi, & Petrosini, 2011; Nicolson, Fawcett, Brookes, & Needle, 2010; Vicari et al., 2005) and **language** (Berninger, 2000; Joanisse, Manis, Keating, & Seidenberg, 2000; Koster et al., 2005). The studies reported here represent just a small anthology of the literature; the actual number of deficits proposed by scientists is much larger.

Areas of inquiry that have received much interest in the past decades due to their perceived potential to shed more light on the underlying causes of the observed impairments are the genes and brains of dyslexic readers. However, as mentioned above, the current empirical record of neurobiological facts and neural and genetic correlates of behaviour related to developmental dyslexia has not been decisive in the resolution of *any* theoretical dispute that existed before in vivo brain-imaging and gene-sequencing became available as tools for scientists. It has not provided a consensus on the aetiology of the reading impairment (Dowker, 2006; Marinelli, Angelelli, Di Filippo, & Zoccolotti, 2011; Pugh et al., 2001; Ramus, White, & Frith, 2006; Ramus, 2003a, 2004; Stanovich, 1985, 1988), neurobiological evidence has not given more scientific credibility to one theory of (impaired) reading and spelling over others (Heath, Bishop, Hogben, & Roach, 2006; Howes, Bigler, Burlingame, & Lawson, 2003; Mechelli, Gorno-Tempini, & Price, 2003; Ramus & Szenkovits, 2008; Ramus et al., 2003) and it has certainly not helped with achieving a clearer definition of developmental dyslexia as a specific learning disability or developmental psychopathology (Bishop & Snowling, 2004; Fletcher & Lyon, 2008; Frith, 1999; Landerl & Wimmer, 2000; Lyon, 1995; Lyon, S. E. Shaywitz, & B. A. Shaywitz, 2003; Wolf, 1999). In fact, new disputes about the "real" neural correlates of impaired reading have erupted, about the myth of the existence of a visual word-form area (Price & Devlin, 2003), the significance of the observed cerebellar dysfunction (Nicolson & Fawcett, 2006) and how letters and speech sounds are integrated in the brain (Blau et al., 2010; Blomert, 2010). The main difference with the times of Galaburda and Kemper, who in 1979 presented some of the first evidence for anomalous neural organisation in a post-mortem study of the brain of a dyslexic reader, appears to be that now each hypothesised deficit comes with its own neural correlates. Thus providing an apparent existence-proof for the deficit in question. Not surprisingly, most suggested deficits have their own treatment program (see section 1.2.3) and new intervention studies emerge based on studies of brain activity as well (Breteler, Arns, Peters, Giepmans, & Verhoeven, 2010). It is even the case for some hypothesised deficits to have their own associated

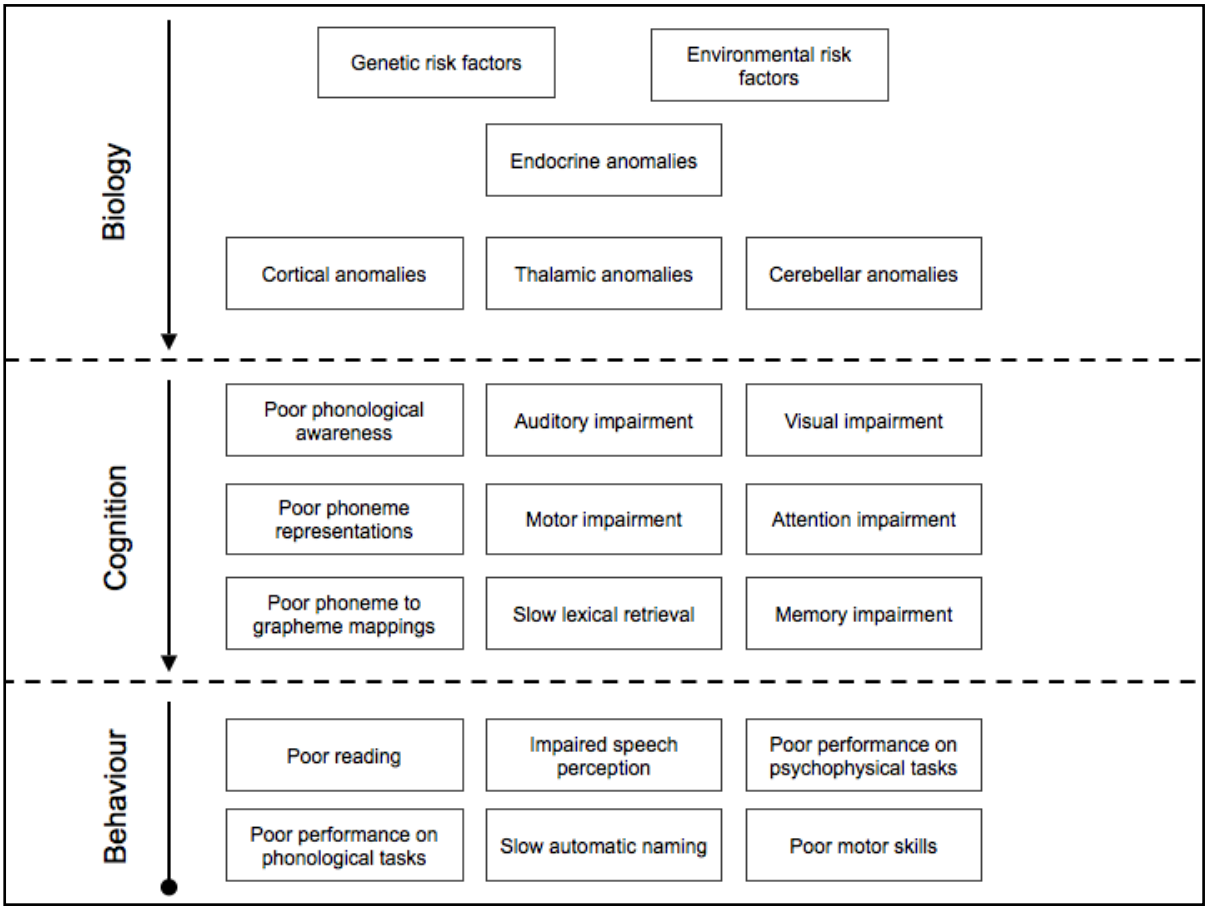

Figure 1.1. The quest for the Holy Grail in the scientific study of developmental dyslexia: Connect the dots and find the proper arrangement of components from biology to behaviour that cause developmental dyslexia.

genetic correlates (Grigorenko, 2001; Ramus, 2004).

How can the difficulties with the acquisition of proficient reading and spelling ability give rise to such diversity in apparently veridical ætiologies proposed by scientists? That is, explanations of causes of the impairment all seem to be evidenced by a considerable empirical record spanning every level of analysis from genes to overt behaviour. Fletcher (2009) interprets the current status quo as the result of the evolution of developmental dyslexia as a scientific concept. After decades of scientific studies some authors are indeed questioning whether the proposed deficits are 'real' (Moores, 2004; Ramus & Szenkovits, 2008) or whether dyslexia actually exists as a collection of deficits (Elliott & Gibbs, 2008). In any case, history may be repeating itself, since summarising papers with a title along the lines of "What have we learned so far?" can be found in almost every decade (Hudson, High, & Al Otaiba, 2007; Snowling, 1996; Stanovich, 1985; Vellutino & Scanlon, 1998; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Most scholars and research programs are well aware of all the other theories out there, but they still embark on a quest for the Holy Grail: To identify a single, one-way, causal pathway from genes to impaired reading. The type of causality the scientists pursue in their quest is a chain of efficient causes hypothesised to be attributable to components or component processes at various levels of analysis as depicted in Figure 1.1. An example of such a quest is a series of papers by Ramus (2003b, 2003c, 2004), who concluded after an extensive review of the literature that a whole range of proposed deficits may be associated with developmental dyslexia, but play no causal role in

its ætiology. In other words, those deficits should be considered epiphenomena of the 'true' cause.

Ramus' quest in that series of articles was to draw causal pathways through the phenomena (the boxes in Figure 1.1) that were substantially backed by empirical results. It is exactly this type of scientific inference, in combination with the inconclusive results of such inferences that –from a meta-theoretical[2] point of view– raises the question whether the proposed causal mechanisms should be considered similar (high *similitude* of theories, see paragraph 1.6). Perhaps they have a low 'truth-likeness' (low *verisimilitude*), or, they are not appraised in a rigorous fashion in order to assess their degree of 'trueness'. I take that it is uncontroversial to suggest that the purpose of a science is to evaluate whether the claims it produces about the way the world works have some truth to them and eventually select the most truthful of all such claims[3]. In the remainder of this chapter I will attempt to analyse the origins of the theoretical diversity and the apparent incapacity of the scientific method to resolve it by means of appraising verisimilitude of individual claims. The conjecture I have made my goal to elucidate in this chapter is: *A theoretical account about the ætiology of developmental dyslexia of higher verisimilitude than current accounts, should be able to explain how it is possible that the current plenitude of theories can be hypothetically true at the same point in time* (or space for that matter, see the recent special issue of Dyslexia 2013, volume 16, issues 3 and 4: Investigating the links between Neurocognitive Functions and Dyslexia).

It is thus explicitly not the purpose of this chapter, or any other chapter in this book, to describe a successful scientific quest that ends in the discovery of the Holy Grail: A unique causal pathway. An empirical quest will be reported in the chapters that follow, but its Holy Grail takes on a very different shape and form. The results of the (meta-)theoretical analysis and historical and philosophical perspectives on the conjecture-made-goal described above, have motivated virtually all of the theoretical and empirical decisions made in studies described in chapters 2-5.[4] The present chapter is an essential general introduction to the more specific empirical inquiries into the role of speech perception in developmental dyslexia.

1.2 Corroborative evidence: A growing body, or morbid obesitas?

The introduction to the scientific study of developmental dyslexia so far, requires from me to evidence at least two claims and offer some explanations if those claims turn out to be true: *Claim 1* - The number of substantive[5] theoretical accounts for the ætiology of

[2] Meta-theory is the empirical study of scientific theorising (Meehl, 1992; 2002; 2004a).

[3] I will use the term 'truth' here in the most practical sense possible: To distinguish between scientific claims, that is, the precision and accuracy with which they describe observable phenomena in some domain of reality. If 'truth' is an uncomfortable word in this context, I see no objection to exchange it for 'scientific credibility'. There is however a formal distinction between a *hypothetico-deductive* test and a *crediballistically-deductive* test of scientific claims (see Rozeboom, 1982).

[4] Pardon my hindsight bias. Chronologically, this chapter was completed last, but note that five years ago, I could not have written it.

[5] I will use 'substantive theory' almost exclusively to refer to a theory backed by a substantial amount or quantity of empirical evidence. This is somewhat different from what Paul Meehl described as "The theory has money in the bank". The latter refers to the track record of corroboration events that constitute risky predictions by a theory and it is questionable whether the predictions by the theories I will discuss can be categorised as risky. Or corroborated for that matter.

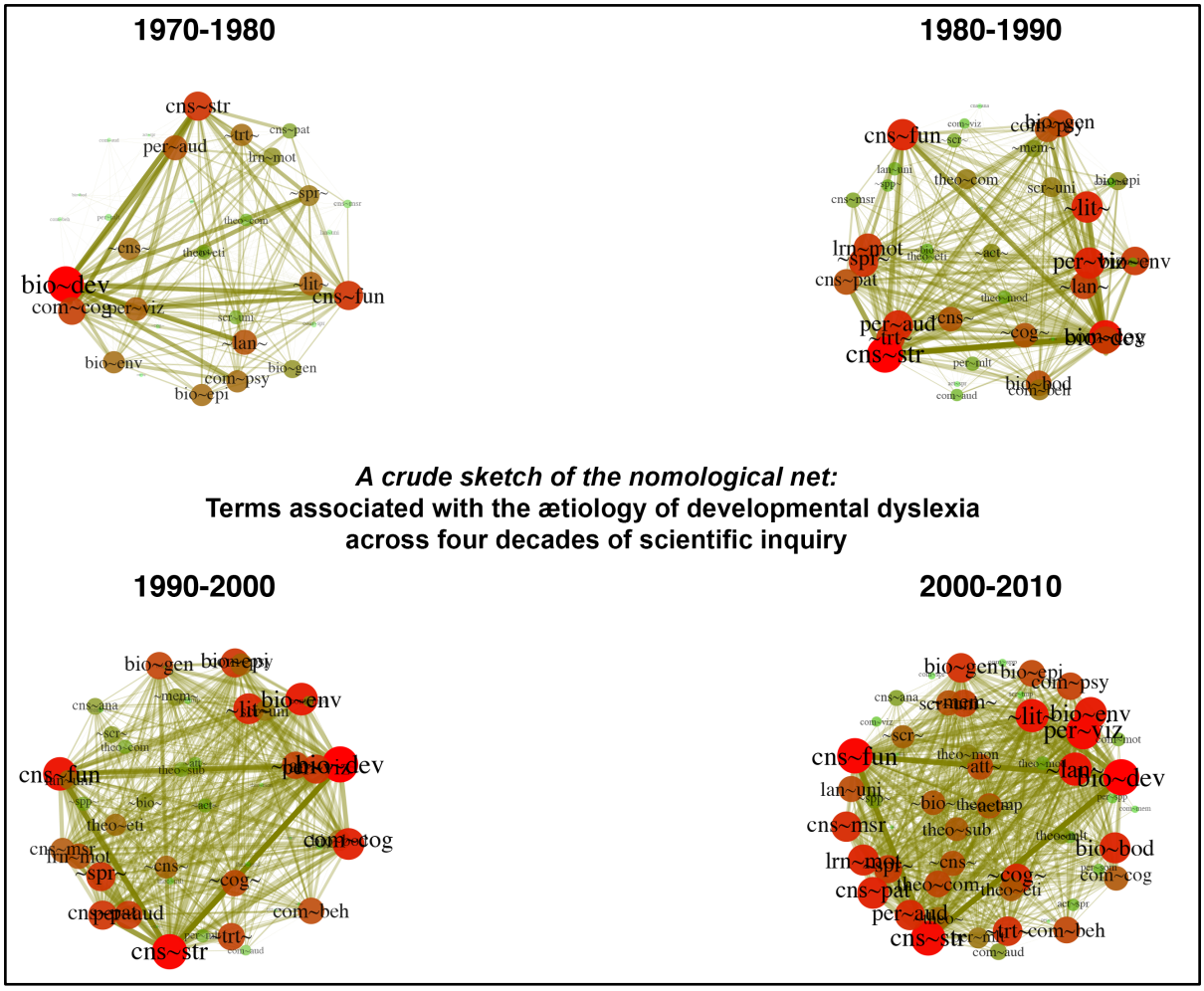

Figure 1.2. Weighted graphs of term associations used in four decades of the scientific study of the causes of developmental dyslexia. The associations represent a cumulative sum of terms occurring in the same abstract of a scientific paper (*N* = 1407). The abstracts were obtained via a PubMed search query on the ætiology of developmental dyslexia. The large, red nodes have a high (standardised) degree, that is, more connections to other nodes in the network (see text and Appendix A.1 for details).

developmental dyslexia has been growing rather than shrinking. *Claim 2* - High quality data measured to corroborate one theoretical account and its causal pathway is of no consequence for the perceived verisimilitude of a competing theory.[6] To do so I present a historical analysis of the terminology used in the scientific literature on the ætiology of developmental dyslexia.

What exactly causes developmental dyslexia has been debated for quite some time, at least since competing theories appeared in the literature in the 1970s. These theories disputed whether the reading deficit was caused by impaired auditory processing or impaired phonological processing (Bradley & Bryant, 1978; Tallal & Piercy, 1973) and were further developed based on evidence from post-mortem studies of the brains of dyslexic readers (Galaburda & Kemper, 1979; Galaburda, LoTurco, Ramus, Fitch, & Rosen, 2006). Both assumed an impaired theoretical entity, the phoneme representation, to be causally entailed

[6] Stated less formally, there appear to be no theories with silverback-alpha-male-pack-leader-results, to which theories competing for resources and proliferation succumb (*empirical grid-lock*). Or, perhaps there are only alpha theories; *all Indians, no chiefs*, and this is a stand-off (*ontological indifference*).

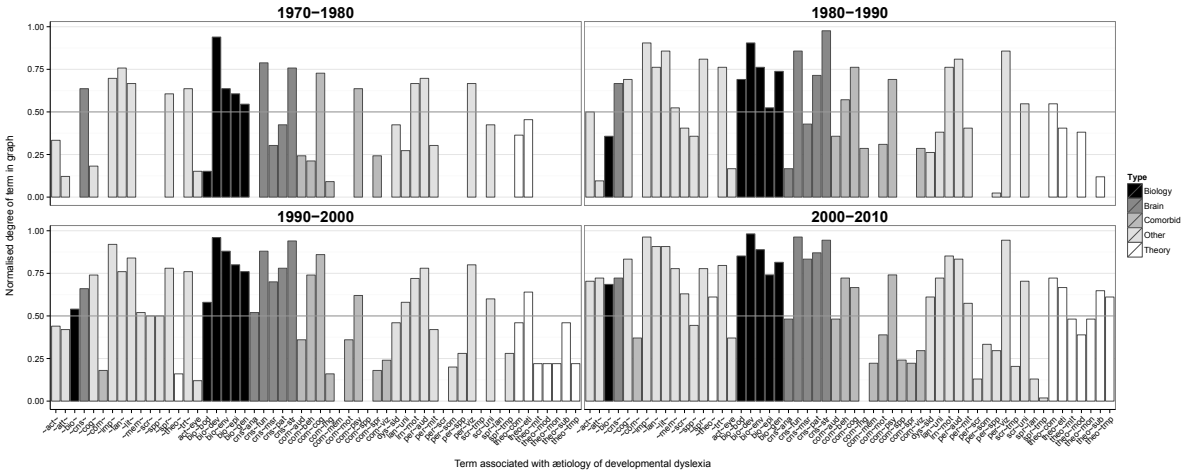

Figure 1.3. Bars represent the normalised degree distribution for each node in the graphs of Figure 1.2. This represents the number of connections each term has to other terms divided by the total number of connections in the graph. The bars are colour coded to indicate terms that can grossly be categorised as Biology, Brain, Comorbidity and Theory (compare to levels of Figure 1.1). See text and Appendix A.1 for details, such as, a list of the terms and their meaning in Table A.1.

in causing the difficulties with acquiring a proficient level of reading and spelling. What may be reasonably expected of a discipline of science is that it works towards a more unified formal description of reality instead of generating more contradictory ones. Figure 1.2 reveals, among other things, that the latter situation seems likely when scientific descriptions of the ætiology of developmental dyslexia are concerned.

The graphs in the figure are representations of the connectivity between terms (categories of words) found in 1407 abstracts of scientific papers on the ætiology of developmental dyslexia on PubMed (see Appendix A.1 for details). The abstracts were analysed as a corpus based on the decade of their publication date. As an example, take the 1970-1980 era. The boundaries of the graph are inhabited by a few terms that are closely associated. The upper side contains auditory perception (`per~aud`), pathology in brain structure (`cns~str, cns~pat`), treatment (`~trt~`), motor learning and speech production (`lrn~mot, ~spr~`). These terms and associations seem to evidence studies about the impaired auditory processing hypothesis mentioned earlier (Tallal & Piercy, 1973). The other two clusters contain terms like language and literacy (`~lan~, ~lit~`) associated with brain function (`cns~fun`) and this seems in accordance with the impaired language processing account mentioned earlier (Bradley & Bryant, 1978). There are also many connections to comorbid diagnoses, cognitive or psychopathological, associated with biological terms including development (`com~cog, com~psy, bio~dev, bio~env`) and visual perception (`per~viz`). No doubt these terms appear due to attempts to define and diagnose dyslexia, but they also herald the arrival of the (visual) magnocellular deficit hypothesis (cf. Stein, 2001).

The graphs reveal that the structure of the graphs does not change just due to an increasing number of terms (the nodes or vertices), but it is clearly the case that the terms become less uniquely specified. That is, the number of connections (edges) a node has to other nodes in the network increases and nodes representing new terms do not lag behind 'old' ones in this respect. It really seems to be the case that what started with roughly two to three substantive theoretical accounts has grown into a vast collection of different theories, deficits and hypotheses, all backed by a considerable empirical record that contains every

type of data from behavioural, neurophysiological, to genetic. Also, every design seems to be represented, there are data sets that were acquired experimentally, by means of decade long prospective studies and/or by comparing subpopulations, as well as computer simulations. It is truly an impressive body of evidence, but what does it evidence?

The pattern inferred from eyeballing the term networks can be quantified by calculating the degree distribution in the network. Figure 1.3 represents for each term the normalised degree of the node. A normalised degree of 0.5 (the horizontal line) means that the connections of a node to other nodes consist of 50% of the unique connections that are available in the network. This does not imply that a node is particularly important; in a fully connected network all the nodes host 100% of the connections. Interestingly, more varieties of *comorbid diagnoses* are mentioned in association with other terms as time goes by. This could be related to the observation in the first paragraph that a definition of developmental dyslexia is still mostly a definition of exclusion criteria, like excluding comorbid diagnoses. Another pattern that emerges is that as more terms appear, they are often connected to more than 50% of the other terms in the network. Terms categorised as *Biology* (e.g., genes, epigenetics) and *Brain* seem to 'lead the way' in reaching the 50% mark, they are almost always mentioned in association with any of the other terms in the network.

1.2.1 The nomological network: A rough sketch

To interpret what the degree signifies, it is important to consider what kind of network this is. What do its nodes and edges represent? I believe it is sensible to argue that the networks represent a very crude sketch of a nomological net as suggested by Crohnbach and Meehl (1955). They described nodes of a nomological net representing theoretical (or ontological) entities and their connections to other entities as lawful relations (functional or compositional). Using the terminology of graph theory and complex networks in combination with basic meta-theoretical concepts, I suggest a more detailed specification of the nomological net is possible. The formalised theories of physics will often consist of deduced entities and laws, whereas theoretical accounts in social science are predominantly based on induction by statistical regularities (see Table 1.1 for a reminder of the differences between deduction and induction, cf. Salmon, 1999). If laws were allowed to be either universal or statistical and the entities can pertain to particular facts or general regularities, the four types of scientific explanation in the received view of scientific explanation (Hempel & Oppenheim, 1948; Hempel, 1968) could be represented by the net (see Table 1.2). By using a weighted and directed graph, one could indicate Deductive-Nomological / Deductive-Statistical explanations as directed edges (one way connections) and Inductive-Statistical explanations by equating the degree of inductive strength to the edge weight.

An ideal rendering of a nomological net should thus be able to represent a topology of scientific explanation: A single theory, a larger theoretical framework, or all the different theories that can be connected in some way. Only a subset of the net will make contact with the empirical record, that is. The verisimilitude of purely theoretical entities could be the node degree or if observable entities are associated to it statistically, the sum of the edge weights could be used. The structure of the topology can be thought of to represent the

Table 1.1

The Differences Between Deduction and Induction. Adapted from Salmon (1999, p.11)

DEDUCTION	**INDUCTION**
All humans are mortal *Socrates is human* **Socrates is mortal**	*All observed ravens have been black* **All ravens are black**
1. In a valid deductive argument, all of the content of the conclusion is present, at least implicitly, in the premises. Deduction is *non-ampliative*.	1. Induction is *ampliative*. The conclusion of an inductive argument has content that goes beyond the content of its premises.
2. If the premises are true, the conclusion must be true. Valid deduction is *necessarily truth preserving*.	2. A correct inductive argument may have true premises and a false conclusion. Induction is *not necessarily truth preserving*.
3. If new premises are added to a valid deductive argument (and none of the original premises is changed or deleted) the argument remains valid. Deduction is *erosion-proof*	3. New premises may completely undermine a strong inductive argument. Induction is *not erosion-proof*
4. Deductive validity is an *all-or-nothing* matter; validity does not come in degrees. An argument is totally valid or it is invalid.	4. Inductive arguments come in different *degrees of strength*. In some inductions the premises support the conclusions more strongly than in others.

'success' of scientific theorising, or its global verisimilitude[7]. The goal is to strengthen the logical structure of the net by strengthening connections in a derivation chain upon corroboration or deduction, but pruning derivation chains that fail severe testing are logically inconsistent. Graph theory and network analysis also allow detecting similitude of graphs (isomorphism) this offers a potential tool to examine whether unification is possible.

The networks under consideration here do not represent this ideal rendering of the net. The nodes could represent theoretical constructs and perhaps the edges even laws, but not as just described. First, it is important to remember that the terms in the networks are conditional on the result of a search query that contained terms used to refer to developmental dyslexia and its ætiology (otherwise a much larger set of abstracts would have been found). The nodes representing the search terms (dyslexia, etc.) have been deleted, as we already know they appear in each abstract and are therefore connected to all the other terms. Otherwise each hub in the network would be connected to nearly all other hubs due to these terms. Second, the terms that make up the nodes are general categories, not strictly representing theoretical entities as some words just mean different things in different contexts. Mentioning a brain structure in an abstract could have theoretical reasons as well as be the result of an empirical study of that structure. Finally, the weight of the edges connecting the nodes in the graphs is calculated as a sum over the abstracts, indicating how often the term pairs were mentioned in concert in the corpus. These weights could very well evidence an underlying empirical law, as the terms must be correlated in some way, but this cannot be interpreted as a correlational measure. As for the structure of the network topology, I see no reason why the general idea of the structure of scientific explanation represented as entities that can be related either by frequency of co-occurrence or by the number of connections they make to other terms does not apply here. This is in

[7] Yes, objections were raised against the received view, most importantly that the goal of scientific explanation is unification, not causality (Kitcher, 1989). In the everyday practice of empirical science, causality will be on the mind of the researcher, not unification. The received view seems to me a most complete account of scientific explanation that strikes a balance between idealised science and a taxonomy of theorising in the wild. Moreover, unification is explicitly defined as a goal of science in the notion of strengthening the logical structure of the net. Also, the received view has been declared a straw man of philosophy of science and has been defended quite successfully recently (see Lutz, 2012).

Table 1.2

The Four Forms of Scientific Explanation According to the Received View. Adapted from Salmon (1999, p.16). Note that the Nomological Net Metaphor implies the Goal of Unification.

	Explananda	
Laws	*Particular Facts (Ideographic)*	*General Regularities (Nomothetic)*
Universal laws	D-N Deductive-Nomological	D-N Deductive-Nomological
Statistical laws	I-S Inductive-Statistical	D-S Deductive-Statistical

fact how graph theory was used recently to analyse the structure of symptoms of psychopathology listed in the DSM-IV (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011).

Even when these restrictions on the scope of the results are taken into account, from the perspective of theory evaluation as an attempt to strengthen those structures in the net that are scientifically credible, and at the same time prune the less credible laws and their associated nodes, the graphs do not show a tendency towards less connected nodes, the opposite is the case. When each term ends up associated to every other term, the network cannot be very informative as a knowledge base. Lebel and Peters (2011), discussing claims of evidence for paranormal phenomena (Bem, 2011), suggested that a weak knowledge system contributes to the publication of such incredible claims in psychological science. The verisimilitude of the entities and laws involved in a theoretical claim should be severely tested and ruthlessly discarded if they cannot pass the tests. That is how the logical structure of knowledge represented in the nomological net increases. In addition to (or instead of) a possible low logical structure of the theories represented by the networks in Figure 1.2, one could interpret the observed non-specificity as the sign of a high similitude of the theories that spawned the studies whose abstracts were analysed.

This could be an explanation for *Claim 2*: Empirical evidence cannot be decisive in assessing the verisimilitude of competing theoretical claims, not because they are not severely tested, but because the entities and laws from which observations are predicted are essentially two sides of the same structural coin. The question of similitude will be discussed in the last paragraph of this chapter.

1.2.2 Meet ~theo~ … there goes the neighbourhood!

The final support for *Claim 1* (the number of substantive theoretical accounts has been growing rather than shrinking) and its logical consequence *Claim 2* (the credibility of substantive theoretical accounts is invulnerable to apparent corroborations of competing substantive theoretical accounts) will be provided by examining the category *Theory*. The bars representing different theories are grouped at the end of the x-axes in Figure 1.3. These terms exclusively represent word combinations associated with a particular theoretical account, from multi-causal theories to those that focus on deficits in temporal processing (see Table A.1). One could object that their appearance is normal due to the progress of a scientific field of study and advancement of technology in general. The number of studies with downloadable abstracts published on the subject seems to grow as a power law of 2^n (n = 36, 123, 264, 686 respectively). The networks do reveal that more detailed, specific terms appear in the scientific record as time goes by. One could argue that as science advances and

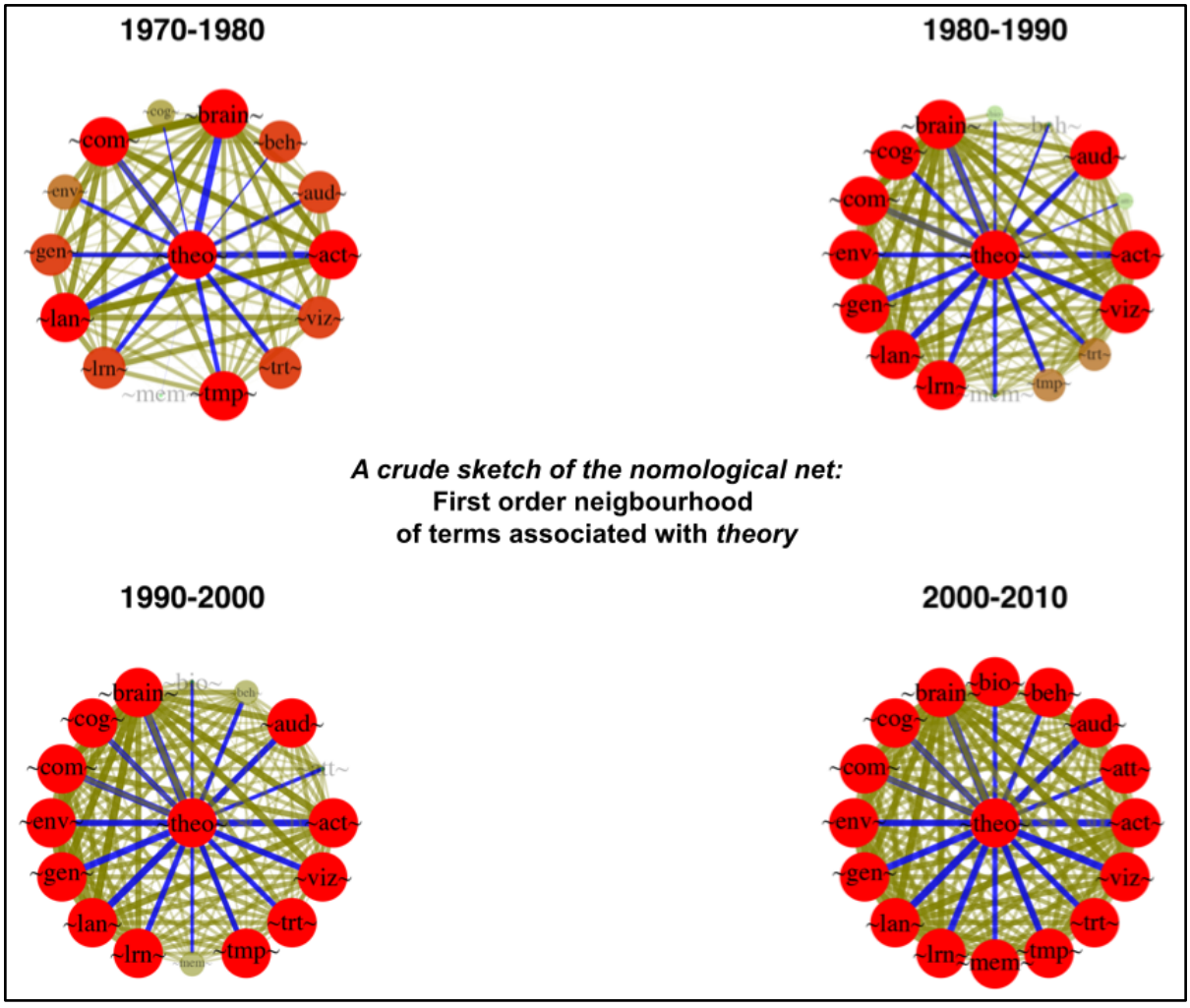

Figure 1.4. The graphs represent the first order neighbourhood of the `~theo~` node based on the networks displayed in Figure 1.2. Here, specific classifications of terms were collapsed into broader categories. The nodes in the graphs now have an equal number of categories for each decade. The blue connections represent direct associations of theory terms to other categories. The other connections represent associations between the terms in the first order neighbourhood of node `~theo~`. See text and Appendix A.1 for details.

advanced technologies like in vivo brain imaging and behavioural-genetic analysis are more commonly available, this will boost connections between terms that were previously unconnected; a logical consequence of scientific progress. To gain more insight in the relevance of the evolution of this nomological net for appraising and amending theories about the ætiology of developmental dyslexia, the corpus of abstracts was re-analysed as shown in Figure 1.4. The focus of the analysis is to assess the impact of the explosive growth of the empirical record on theory evaluation. To do so, all the specific terms were collapsed into 16 broader categories that occurred in all four decades and every word indicating a theoretical statement was collapsed to the term `~theo~` (17 nodes, see Appendix A.2, Table A.2). As a next step the node representing theory terms and its direct neighbourhood was isolated from the total network. This is a neighbourhood of order 1, meaning it contains only nodes that have one direct connection to the theory node. Those connections are displayed in blue in Figure 1.4, the other connections are those that exist between the other nodes in the neighbourhood.

Using the broader categories, the networks now appear to present a more balanced picture. In 1970-1980 on this much coarsely grained scale, terms directly associated to theory indicator words were also connected to each other. Important is to note the difference in the weights of the connections (how often did the terms appear together in an abstract in that decade?) and the degree of the nodes (how many connections does the node make to other nodes?). These values (represented by size and colour of nodes and edges) are rescaled to the minimum and maximum values observed in each decade and therefore comparable across decades. The conclusion must therefore be that the pattern persists: The degree distribution and weights of the connections of the nodes in these neighbourhood networks increase over time. One way to express this for an individual node is to look at its co-citation coupling (Small, 1973). Two connected nodes are co-cited if another node 'cites' (connects to) both of them. For each of the four decades the average co-citation coupling of `~theo~` to the other nodes is: 11.7 (*SD* = 1.9), 14.3 (*SD* = 1.1), 14.6 (*SD* = 0.7) and 15 (*SD* = 0), respectively. This means that in the last network where `~theo~` is connected to `~gen~` *all* the other nodes connect to *both* `~theo~` *and* `~gen~`. Such is the case for all possible node combinations with `~theo~` as there is no dispersion in co-citations. A summary metric of the weights of the connections is usually expressed as the sum of all edge weights connecting to a node, or the graph strength (Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004). For the `~theo~` node the graph strength is: 24.4, 41.2, 52.7 and 69.0 for each consecutive decade.

To summarise: More connections emerge between theory words and other terms as time goes by and the strength of those connections (how often they co-occur in abstracts of scientific articles) also increases. In the most recent decade (2000-2010) the terms found in 686 abstracts that are most directly associated with theoretical claims about the ætiology of developmental dyslexia form a network in which:

1) Indicators of theoretical claims are associated to every other term in the network.
2) The strength of these associations is equal and maximal for all connections.
3) Each node that is associated to a theory word is also connected to every other node in the network.

1.2.3 The 'inventors' in social science: Interventions as technology.

What are the causes and consequences of this weak knowledge system? One consequence that should raise some concern is the development of 'technology' based on these ætiologies of developmental dyslexia. In the "applied" social sciences the scientific knowledge produced by the 'pure' fields is turned either into diagnostic or performance measuring instruments or into intervention programs whose goal is behavioural change, which can be anything from optimal design of advertising campaigns to a treatment program for a psychopathological disorder.

In the case of developmental dyslexia, the technology produced is intervention programs that ultimately should ameliorate reading and spelling performance. The fact that just about any theoretical claim seems to enjoy empirical success has resulted in an equal plethora of intervention studies, some of which are so esoteric, or 'armchair unlikely', that I consider it a real possibility that as a consequence of the inability of scientists in these fields to properly evaluate their theories for verisimilitude, the lives of children were adversely affected. At the very least it should raise questions about the ethics of intervention research. A whole range of interventions, some without any apparent relation to reading and spelling, continue to be proposed by serious scientists in serious scientific journals. The interventions

reveal we are dealing with a science that is realist (in the most direct sense) about the phenomena and entities its theories posits to exist in reality. In principle, this is not at all a bad trait for an individual scientist, as long as there is the realisation that the ontology is a temporary one. In an arena with so many different competing theories claiming to 'get it right', one would expect some more pervasive scepticism from a genuine scientific endeavour.

The proposed treatments are clearly inspired by the (causal) ontology associated with a specific deficit (see Figure 1.1.) and the rationale behind most of these interventions is not very different from attempting to repair a faulty component in a machine. Recent examples of studies reporting treatment effects are: Adding Fish Oil (Omega-3 fatty acids) to the diet to improve the diminished myelination of the magnocellular part of the central nervous system (e.g. Cyhlarova et al., 2007 -> *magnocellular deficit hypothesis*); Using coloured lenses and coloured overlays to improve reading fluency (e.g. Lightstone, Lightstone, & Wilkins, 1999; Whiteley & Smith, 2001) –> *visual disturbance / visual stress hypothesis*); Music therapy (Cogo-Moreira et al., 2012; e.g. Overy, 2003) –> *rhythm / coordination imbalance hypothesis*); Intensive training with tonal sweeps and acoustically modified speech (e.g. Tallal, 2004 –> *auditory temporal processing hypothesis*); Neuro-Feedback training (e.g. Breteler, Arns, Peters, Giepmans, & Verhoeven, 2010 –> *brain dynamics deficit hypothesis*); Training motor skills (e.g. D. Reynolds, Nicolson, & Hambly, 2003; D. Reynolds & Nicolson, 2007 – *cerebellar deficit hypothesis*); Presenting stimuli to visual, auditory and touch modalities opposite a dysfunctional hemisphere (e.g. Smit-Glaudé, van Strien, Licht, & Bakker, 2005) –> *hemispheric balance model of reading and dyslexia*); Training Rapid Serial Naming of pictures, colours, numbers and letters (e.g. Eleveld, 2005 – *rapid naming / fluency deficit hypothesis*); Training children by letting them play action video games (Franceschini et al., 2013) -> *visuo-spatial crossmodal temporal attentional deficit / (dorsal) magnocellular deficit hypothesis*.

These interventions were chosen because they appear rather unorthodox and they all but two report conclusive positive results, some of which are close to miraculous. Twelve hours of 'training' with action video games was reported to cause gains in reading, spelling and phonological skills equivalent to 1 year of 'regular' remediation (Franceschini et al., 2013). There are many more reports of effective interventions in the literature, most of which are based on such regular remediation programs: Some variation on repetition learning of reading and spelling performance.

As Novalis suggested in the epigraph, the inventors are also the true hypothesisers in social science. They cast their nets and they always catch. This problem of 'easy' evidencing of effects of intervention was recently discussed as one of the pervasive problems in psychological science (Boot, Simons, Stothart, & Stutts, 2013). The that are solutions offered mostly concern the design of experiments (e.g., active control groups as placebo stand-ins) and fail to recognise that interventions are in fact a technology produced by scientific theories and tests of their efficacy should perhaps be treated differently from tests of theoretical entities (see Meehl, 1997). As has been known for a long time, effective technologies of the social and life sciences, from psychotherapeutic or surgical intervention to pharmacological treatments have a high likelihood of 'breaking down' when they are brought into the real world where they have to be efficient as well as mere effective under ideal conditions of a clinical trial (cf. Cochrane, 1972).

My goal, as stated earlier, is to seek an explanation for the persistence of the weak knowledge structure in the empirical record. That is, examine the extent to which substantive theories explaining the same phenomena are similar, and if they are not similar,

the extent to which their simultaneous claim to be 'true' arises due to the weakness of inferential methods used or failure to adhere to a philosophy of natural science. First, I will sketch a meta-theoretical account of the mechanism behind the erosion of authority of theory corroboration by empirical evidence (section 1.3). Second, I will examine the differences between theorising in the empirical social sciences and the natural sciences (section 1.4). Before solutions can be presented, I provide an overview of the problem of theoretical diversity and discuss some causes for its emergence specific for the object of study in psychological science (section 1.5). Section 1.6 is a proposal to define developmental dyslexia as a state-entity and study claims about its ætiology guided by degrees of theory specification. The remaining chapters in this book will be characterised according to the level of theory specification they set out to test.

1.3 Epistemic Sloughing: How to Erode the Authority of Corroborative Events

> "If Psychology is ever to become anything more than a mere aggregation of opinions, it can only be by the establishment of some datum universally agreed to."
>
> – Herbert Spencer (1855, p. 8)

What appears to have occurred in the study of developmental dyslexia is that *any* significant pattern in the data was interpreted as a phenomenon of significance for the ætiology. Although direct tests of predictions by different theories do exist, the empirical results appear to have no authority on the level of perceived verisimilitude of the competing theories, even if those results pertain to empirical studies of the brain or genes and are therefore indirect discorroborations with respect to competing theories (Ramus, 2003a, 2003b, 2004; Ramus & Szenkovits, 2008; Ramus et al., 2006;). These theories are in *empirical gridlock*, or perhaps due to the shear volume of empirical phenomena that are perceived as being relevant, an *ontological indifference* has evolved among the researchers in the community. The energy spent on attempts to dispose of competing theoretical accounts is apparently perceived as a waste of time and resources by the field; in any case, such attempts do not seem to occur often. The result is a flooding of the empirical record with empirical facts representing unrefuted hypotheses, a condition I call Empiarrhhea[8]. Each phenomenon is "saved" to support a specific theoretical deficit and each deficit ends up with its own anatomical and functional brain anomaly and genetic anomaly. If there is anything evidence from neuroscience and genetics has done, it is an ontological truth status for different deficits that resonates more with a generally accepted world-view, than verisimilitude of a theoretical whole: Such deficits *must* exist as constituents of reality, after all, aren't we determined by our brains and genes according to science? This is why in Figure 1.1 the CNS and genes are at the top of the causal pathways, the origin of the derivative chain, with most (apparent) epistemic weight.

It seems as if the scientific method, the repeated application of the empirical cycle in which theories are tested based on their predictions of observable phenomena has turned into a vicious cycle of consisting only of corroboration events. I use the term corroboration event loosely to indicate that a theoretical prediction (and as such the whole of theoretical claims in the derivational chain) was 'evidenced' or 'confirmed' or 'supported' or 'verified' or

[8] I introduced the term Empiarrhea to indicate the zealous empiricists practice of flooding the empirical record with cute facts that have no impact on the veracity of theoretical claims. Rick Dale (personal communication, 18-03-2011) pointed out one could also call this phenomenon Theorrhea indicating the influx of new theoretical accounts into the scientific record by theorists in search of a monism.

'not falsified' or 'considered plausible'[9] by an empirical observation according to a cultural convention specific to a community of scientists. Often this convention concerns turning the prediction into a hypothesis that is tested using an observational threshold for evidencing phenomena. The threshold represents a certain degree of confidence about the truth-status of the hypothesis, for example, whether the probability of an observation falls below the *α-level* (false positive rate) in null hypothesis significance testing (NHST). The 3σ *evidence level* and 5σ *discovery level* for elementary particles is used to decide how remarkable an observation is, compared with observations that may be expected if the predicted particle was not part of the description of reality (i.e., the standard model).

The corroboration strength, or weight of the event granting truth-likeness to a theory in the perception of scientists, should however not be confused with the magnitude of a probability or an effect size associated with the event. Together with a complex interaction of less objective judgements about the novelty, aesthetics, or risky-ness of the prediction and the severity of the test, it is the *objective precision and accuracy of a theoretical prediction that decide the corroborative strength of the event* and its effect on perceived verisimilitude of the theories involved (Mayo & Spanos, 2006; Meehl, 1990b, 2002). For example, in general a lesser corroboration strength is assigned to a *convergence of evidence* compared to the *prediction of novel facts*. General Relativity Theory (GRT) could account for Mercury's perihelion advance, a known anomaly to Newton's theory of celestial mechanics, but it was the corroboration event of a predicted observation (bending of star light during an eclipse in 1918) that convinced most scientists, and the general public (compare Hasselman, 2013 on using model fit indices as corroboration strength). One group of scientists formed an exception, most astrophysicists were more impressed by the after-the-fact prediction of Mercury's perihelion advance (Brush, 1989). Meehl (2002) offers an interpretation that I believe is crucial for understanding the weak structure of the nomological net described in the previous section. I will refer to it as the *epistemic sloughing*[10] of theoretical predictions by means of hypothesis testing. In order to understand the mechanisms of this process consider the different types of research that are commonly used in the empirical fields of the social and life sciences.

The kind of research conducted in the social and life sciences can be categorised (based on Lykken, 1968) as follows:

1) The effect of some treatment on some output variables in applied research (e.g., the testing of technology, (cf. Meehl, 1997). These studies are a special case of studies that:
2) Examine the difference between two or more groups of individuals with respect to some variable. These studies are again a special case of studies that:
3) Examine the relationship or correlation between two or more variables within some specified population.

In general, it can be said that data patterns that might evidence or refute hypotheses, irrespective of the kind of study that is conducted, will be patterns of association (e.g., variables share a degree of common variance, value sequences display a degree of recurrence or symmetry (breaking); Neyman, 1969). Whether or not an observed

[9] These terms mean different things to philosophers of science and logicians. I hope my generalisation for the sake of the argument will be forgiven.

[10] Admittedly, I had not heard of this word before I read *"Sloughing Ontology"* (Dale, 2008). Given our Theorrhea / Empiarrhea exchange, it seemed appropriate to use Epistemic Sloughing here.

association between variables represents a causal entailment is a matter of interpretation of the results in the context of the study as a whole. At the most basic level, all studies are tests of sign predictions.

To predict the sign of a correlation is the least risky of predictions[11] a theory could produce about future empirical observations. The values can take on: Positive, negative, or no correlation, or in terms of a difference between variables: Larger, smaller, or about equal to 0. The most risky predictions a theory can make are point-value predictions (actual measurement outcomes), because the variation in possible outcomes of such future observations is countably or uncountably infinite. To test a prediction in the social and life sciences, most often a hybrid of Fisherian significance testing and the Neyman-Pearson paradigm of Null Hypothesis Significance Testing (NHST) is used (Nickerson, 2000 is a review of the use of NHST in social science)[12]. Other fields may use different procedures, but generally speaking it involves estimating how likely it is to have observed the data, if the theory would not have urged us to make an observation (the chance of observing the data, or more extreme data, given that the or the null-hypothesis (H_0) is true). This concerns data patterns that could have revealed *–absent the theory–* the predicted observational constraint against the background noise of 'obvious' observations. The predicted observation has to stand out above this background noise and the null-hypothesis test is a decision on whether to accept or reject the observation as being background noise. The difference between corroborating the prediction of a positive correlation by rejecting H_0 and an expected instrument reading of 8 decimal points accuracy by rejecting H_0 is lost in the dichotomous accept-reject outcome (Cumming, 2012; Steiger, 2004). In the accept-reject dichotomy lies the meta-theoretical problem: Assigning a weight to the corroboration of the theory when H_0 is not refuted. The epistemic linkage between the substantive theory **T**, and the hypothesis test of its prediction **H***, has been broken (Meehl, 1997, 2002). The exoskeleton of truth-like scientific knowledge that supports the derivation chain of a theoretical prediction is effectively shed by hypothesis testing and interpreting the significance of a body of empirical evidence compares to picking scabs.

So why were scientists studying cosmology more impressed by GRT's convergence on predicting the orbit of Mercury? Meehl suggests it was the extreme numerical accuracy of the after-the-fact prediction. The bending of light phenomenon could also be derived from Newton's theory of gravity and GRT was indeed much more accurate than those predictions. This was however incomparable to the numerical accuracy of Mercury's predicted orbit. As an example of the odd consequences this epistemic sloughing can have on evaluating theories, consider the evidence for the existence of atoms and molecules. In 1913 Perrin made a famous argument in favour of the postulate that atoms needed to be included in the scientific description of reality by showing 13 qualitatively different ways to calculate Avogadro's number (amount of atoms in a volume of space). One could use the kinetic theory of gases, the electrochemical theory of electrolysis or X-rays to arrive at an estimate. Einstein's 1905 observations on Brownian motion were used and one derivation could be made based on the fact that the sky is predominantly blue during the day (Meehl, 1990a; Salmon, 1978). This argument convinced even the most persistent skeptics (except Ernst

[11] Strictly speaking the undirected Boolean prediction: "there is a difference / correlation" or "there is no difference / correlation", is weaker because it can just take on two values. The difference is lost due to the dichotomous nature of null hypothesis significance testing. The significance threshold is adjusted to represent the difference the two (one-tailed or two-tailed test).

[12] I will not consistently distinguish between the two approaches and refer to the whole enterprise of hypothesis testing as NHST.

Mach apparently, see Meehl, 1990b), because it would be extremely unlikely if these different calculations would converge (by order of magnitude) close to $6.02214129(27)\times10^{23}$ mol^{-1} based on chance alone (Meehl estimates the odds are over a quadrillion to one against). As Poincaré lucidly stated, if 13 ways to count something yield approximately the same number, something must have been counted (Meehl, 1990a).

The 13 estimates were however not exactly the same and this did not surprise any physicist, because the theories used in the derivation of the estimates (the auxiliary theories) were perceived to represent different degrees of verisimilitude. In an interesting twist, Meehl (1990b) explores what would have happened if in those days a statistical test were demanded to test the null-hypothesis "there is no difference between the estimates". The result would be that H_0 should be rejected and as a consequence the atomic/molecular theory should be rejected. No physicist would have even contemplated declaring the theory refuted based on this evidence, because the epistemic relevance of the entire derivational chain of the argument leading to this convergence of estimates was uncanny. A concrete indicator of this relevance is the order of magnitude of the values (10^{23}) that was recovered by all 13 estimations. The epistemic significance of this magnitude is lost if one considers the 13 estimates as a distribution of observations that should represent a true score, but take on different values due to measurement / estimation errors (T = X + E).

To give an example of what physical science has achieved in terms of precision of point-value predictions, consider one of the most precise predictions of a measurement outcome by a scientific theory. It concerns the existence of an anomalous electron magnetic dipole moment (which is anomalous to the Dirac equation) by Quantum Electrodynamics (QED). According to the most accurate calculations the anomaly, when measured in a metal trap with a cylindrical cavity whose resonance structure is known, should have a value of: a_μ^{QED} = 11659180.4(5.1) x 10^{-10} (Aoyama, Hayakawa, Kinoshita, & Nio, 2008; Hagiwara, Martin, Nomura, & Teubner, 2007). The most accurate empirical measurement of the predicted value of the anomaly in the described measurement context is: a_μ^{exp} = 11659208.0(6.3) x 10^{-10} (Bennett et al., 2006). Indeed, the residual difference between theory and measurement is not zero, but the precision and accuracy of the theoretical prediction are unparalleled by any scientific theory (this residual difference can however be understood in Quantum Chromodynamics or QCD). No cats were killed in the process of measuring this accuracy (Schrödinger, 1935).

Such examples of accuracy do not exist in any other discipline of science, moreover note that the theory in fact predicts the entire measurement context in which the phenomenon may be observed to yield the predicted measurement outcomes. It would be the same as having a psychological theory that yielded a formula that could be used to predict numerically the difference between running a test on a computer and filling out a paper and pencil version of the same test. In social science, the prediction of measurement context is implicit, but very real; consider the factorial design of experiments or the observation of a variable corroborating a theoretical entity only at the level of a sample of a certain size. In fact, I propose to consider all actions taken to make sure that "everything else being equal" the evidenced prediction should be attributed to the theory (the ceteris paribus clause) reflect expectations by the experimenter about the measurement context in which the phenomenon should be observable.

In principle, the statistical testing of hypotheses is not a problem as long as sensible scientists uphold the epistemic linkage between theory and tested hypothesis in their evaluation. The nature of the prediction made by the theory plays an extremely important role in this respect. The odds of a quadrillion to one against count as a "damn strange

coincidence" (Salmon, 1984), but these odds are the exclusive result of point-value predictions. Therefore, in physics this will hardly present a problem (note, they often do not use statistical tests at all). In the softer empirical sciences, whose theories do not predict anything beyond the direction of an association between variables, a whole range of obfuscating factors make it extremely difficult for scientists who are genuinely willing to do so, to reconstruct the epistemic linkage in a meaningful way:

> "Thesis: Null hypothesis testing of correlational predictions from weak substantive theories in soft psychology is subject to the influence of ten obfuscating factors whose effects are usually (1) sizeable, (2) opposed, (3) variable, and (4) unknown. The net epistemic effect of these ten obfuscating influences is that the usual research literature review is well-nigh uninterpretable." (Meehl, 1990c, p. 197)

If Meehl's thesis is correct and essential mechanisms for deciding the veracity of theoretical claims in a field of science are sabotaged such that its claims are uninterpretable, it cannot be considered a scientific enterprise that generates increasingly accurate knowledge about the structure of the domain in reality it studies. There may be other goals for genuine scientific endeavours to pursue, but I am considering empirical sciences with just cause: In this scenario we cannot expect the applied fields of the social and life sciences to produce effective and efficient technology (e.g., Ioannidis, 2005; Worrall, 2011). Table 1.3 lists Meehl's obfuscating factors of which many will eventually be discussed one way or another in this book. In this section I will address two categories of factors: 1) Factors that indicate a credibility hurdle is necessary for the empirical fields (1.3.1) and 2) Factors indicative of obscurantist practices concerning the derivation and evaluation of a prediction (1.3.2).

1.3.1 The crud factor and the credibility hurdle

Lykken (1968) estimated that the 'unrelated' molar variables involved in most studies in psychology share 4-5% common variance, meaning, with 0 measurement error a correlation of about .20 can be expected between any one of them. This really depends on the field of inquiry, but it seems that estimates between .15 and .35 are by no means an exaggeration (Lykken, 1968; Meehl, 1990a, 1997). Based on the lower estimate, the expected difference between any group-based averages would be about 0.5 standard deviations. The test against the null hypothesis of "no association" is often a test against a "straw man" null hypothesis (LeBel & Peters, 2011), because it can be known in advance that an assumption of no association at all is false (Bakan, 1966; Bower, 1997; Ferguson & Heene, 2012; Gliner, Vaske, & Morgan, 2001; Meehl, 1967; Nunnally, 1960; Rozeboom, 1960). Therefore, a researcher can maximize his chances to corroborate *any* weak prediction of association between variables, by making sure a large enough number of data points are collected. This "crud factor" (cf. Meehl, 1990c) implies a researcher has a chance of 1 in 4 to evidence an association using a sample size of 100 data points, without even needing a truth-like theory to predict an association (Bakker, van Dijk, & Wicherts, 2012; Ioannidis, 2012; Simonsohn, Nelson, & Simmons, 2013). This is of course not something a researcher deliberately plans to do (one assumes), or a *statistical* error that is committed. It is a genuine problem of theory evaluation, rather specific for the domain of the social and life sciences. The problem does have a practical solution and this involves what may be called raising the credibility hurdle for observed phenomena.

By increasing the number of data points to evidence a pattern, one increases the statistical power of the study (its sensitivity) to detect an association when it is indeed truly present in the data as posited by the theory (a 'true effect'). Suppose a researcher publishes a multi-experiment article that reports multiple corroborations of a predicted association

Table 1.3

Meehl's (1990a, 1990b, 1990c) Obfuscating Factors that render Scientific Claims Uninterpretable in the Context of NHST and Weak Predictions. The last Columns display the Effects of the Factors on the Perceived Verisimilitude of Theories that are in fact Truth-like and Trivial or Likely False.

		Effect on perceived verisimilitude of a theory that is in fact:	
	Factor	*Truth-like*	*Trivial / False*
1.	Loose (nondeductive) derivation chain, making several "obvious" inferential steps requiring unstated premises (intuitive, common-sensical, or clinical experience).	-	
2.	Problematic auxiliary theories, although explicitly stated.	-	
3.	Problematic ceteris paribus clause.	-	
4.	Imperfect realization of particulars (experimenter mistakes in manipulation) or experimenter bias in making or recording observations.	-	
5.	Inadequate statistical power to detect real differences at the conventional significance level.	-	
6.	Crud factor: In social science everything correlates with everything to some extent, due to complex and obscure causal influences.		+
7.	Pilot studies used to (a) decide whether "an effect exists" and (b) choose a sample size of adequate statistical power if the pilot effect is borderline but in the 'right' direction."		+
8.	Selective bias in favor of submitting reports refuting the null hypothesis.		+
9.	Selective bias by referees and editors in accepting papers refuting the null hypothesis.		+
10.	Detached validation claim for psychometric instruments.	-	+

between two or more variables. Why publish multiple experiments if the association was evidenced in the first experiment? One reason is purely meta-theoretical; to increase the credibility of the theory that predicted the phenomenon, add to its perceived verisimilitude (in empirical reports of psychological science, authors often report they attempt to show a certain effect is "real"). Assume H_0 was refuted in 5 independent replications at $p < .05$, this would be an impressive corroborative track-record for the theory that predicted the association. That is, if there ever were a truth-like theory that made a prediction. As explained above, this scenario could happen just as likely for any pair of "unrelated" variables, provided the sample sizes are large enough. Knowledge of such facts does not prevent the zealous researcher from the impression that the corroborations of his predictions are really due to the verisimilitude of his theorising (not to forget the ingenuity of his experimental methods of course).

It is possible to solve this problem by taking seriously the rules of statistical inference and the epistemic sloughing effect of the null-hypothesis test when corroborating weak theoretical claims. In the multi-experiment example, the probability of each subsequent rejection of H_0 being a false positive (rejecting H_0 when it is true, a type-I error) represented by the α-level, effectively reduces to $.05^5 = .000000312$ (Schimmack, 2012). This is approximately the "discovery level" used to evidence phenomena in particle physics, like the Higgs Boson (the 5σ criterion). At first sight, this drop in the probability of rejoicing over a false positive seems rather welcome, but what about the probability of committing type II errors (false negatives)? The decrease of the effective α-level represents the fact that the

Table 1.4

*The Power Failure (Decline of Total Power) for Subsequent Corroborations Using the same Sample Size. Values calculated for 3 Different Levels of Cohen's d, using G*Power 3*

N corroborations	Total Power	N for Large	N for Moderate	N for Small
1	80 – 81	52	128	788
2	71 – 72	52	128	788
5	0.5 – 0.9	52	128	788
10	0	52	128	788

distribution of values that could be observed if H_0 were true (the background noise) changes and becomes narrower. This means, everything else being equal, there is a drop in statistical power, sensitivity to detect the true effect with the same sample size.

To maintain equal power across subsequent observations, one needs to increase the sensitivity of the study for each observation that adds to the credibility of the phenomenon. In social science this often means increasing the sample size of the study, in physics one resorts to building a more sensitive measurement apparatus. Table 1.4 shows what happens to the Total Power (of the replications as a whole) if the sensitivity is not adjusted. At 5 subsequent corroborations the Total Power has already dropped below 1%. If H_0 is rejected at this sensitivity level, it is very unlikely the results represent a true effect. This is a very real problem, it was recently concluded by studying a large number of meta-analyses that individual studies in Neuroscience are severely underpowered ranging from 8% to 21% (Button et al., 2013).

Schimmack (2012) provides a table with requirements for a multiple-experiment study to maintain a sensitivity of 80% for each individual rejection of H_0 (Table 1.5 is an excerpt). In other words, in order to be credible as a streak of corroborative events of a theoretical prediction, the observational hurdle has to be increased. The observation of *n +1* significant effects out of *N* attempts at the Total Power level, is unlikely to be due to chance alone (or the crud factor): The probability of observing 5 significant results in 5 studies whose Total Power is 50% is 0.0313. So in just 3 out of 100 five-experiment studies of the same Total Power, we would expect to see 5 significant results. That would probably qualify as a "damn strange coincidence" (Salmon, 1984) if it had occurred absent the theory predicting the observation. Schimmack (2012) calculates an incredibility index (IC-index) as the binomial probability of observing at least one non-significant result in the streak of corroboration events, given an estimate of the Total Power to detect the effect (post-hoc observed power). For this 5-study example, given a large effect size (Table 1.4) the IC-index would simply be 1-Total Power = 96.9%. That's how incredible these results would be, given the sensitivity of the test. Should I have increased the observational hurdle as indicated by Table 1.5, the IC-index would have been 4.4% and I would have been much more comfortable to proclaim to have evidenced an association predicted by a theory.

There is no tradition to raise the credibility hurdle for subsequent corroborations of associated variables in the social sciences and in combination with the crud factor a dangerous recipe for immobilising the scientific method emerges. Suppose dyslexic and average readers in fact belong to the same population of normal variation of reading ability and the only "true" characteristic that separates these groups is some demarcation of ability in the lower end of the reading ability distribution (see Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992). Just by taking into account a crud factor, it may be expected that reading

Table 1.5

*The Relation between Subsequent Corroborations Events by Statistical Significance and Sample Size needed to maintain 80% Power for each individual Corroboration. Values calculated for 3 different Levels of Correlation |r| (based on Cohen's d), using G*Power 3.*

N corroborations	Total Power needed	Large (|*r*|=.4)	Moderate (|*r*|=.25)	Small (|*r*|=.1)
1	80.0	52	128	788
2	89.4	136	336	2068
5	95.6	440	1090	6750
10	97.8	1020	2560	15820

ability correlates with any other variable, meaning, if we sample a group based on low reading ability, this group can be expected to have low scores on any other variable as well. The structure of the networks in section 1.2 seems to be a logical result of this ambient correlation. To illustrate the consequences of the premise, consider a numerical example by Meehl (1990a; see also Meehl, 1990c). Imagine he is talking about reading ability (x) and some random other variable, for instance, amount of music education received during childhood (y):

> "I provide one simple numerical example to illustrate the point that a modest crud factor cannot be discounted in the metatheory of significance testing. […], suppose that a representative value of the crud factor in a certain research domain were r = .30, not an implausible value from the examples given. We have a substantive theory T, and we are going to "test" that theory by a correlational study involving observable variables x and y, which, however, have no intrinsic logical connection with T and have been drawn randomly from our huge pot of observables. Assume both x and y are approximately normal in distribution. We dichotomize the independent variable x at its mean, classify each subject as high or low on the x trait, and compare their scores on the dependent variable y by a t test. With the mean standard score of the highs on x being .8 (at + 1 MD) and that of the lows being -.8, there is a difference of 1.6 sigma in their means. Hence the expected mean difference on the output variable is d = .48, about half a sigma. Assuming sample sizes for the highs and lows are around 37 (typical of research in the soft areas of psychology), we find that the probability of reaching the 5% level in a directional test is .66. So a theory that has negligible verisimilitude, and where there is no logical connection between the theory and the facts, has approximately a 2-to-1 chance of being corroborated provided that we were predicting the correct direction. If one assumes that the direction is completely chance (which in any real research context it would not be, for a variety of reasons), we still have a .33 probability of squeaking through with a significant result; that is, the empirical probability of getting a positive result for the theory is larger, by a factor of 6 or 7, than the .05 we have in our minds when we do a t test" (Meehl, 1990a, p. 125)

In this example the high/low cut is made at the mean of x, but if this were a study in developmental dyslexia and x were reading ability, the cut would be much more extreme (10^{th} or 25^{th} percentile). The probability of finding a positive result without a theory is likely somewhat higher than 33% for a non-directional hypothesis, because in studies of developmental dyslexia one is always comparing whether the difference in means of performance measures is greater (or smaller) than 0.

Whether this complicates matters can be relatively easily tested. Figure 1.5 shows results that were obtained from a simulation of different levels of the crud factor and dyslexic reading selection criteria. First, three different population-level correlations values were used to simulate the crud factor association between any two random variables x and y (N_{pop} = 500000, with a correlation of x and y of .1, .2 and .3). Second, based on x (e.g., a standardised reading ability test), the population was cut into regions to sample dyslexic readers from (the 10^{th} percentile and lower, and the 25^{th} percentile and lower) and average readers (between the 25^{th} and 75^{th} percentile). These cut-offs were used to draw samples to

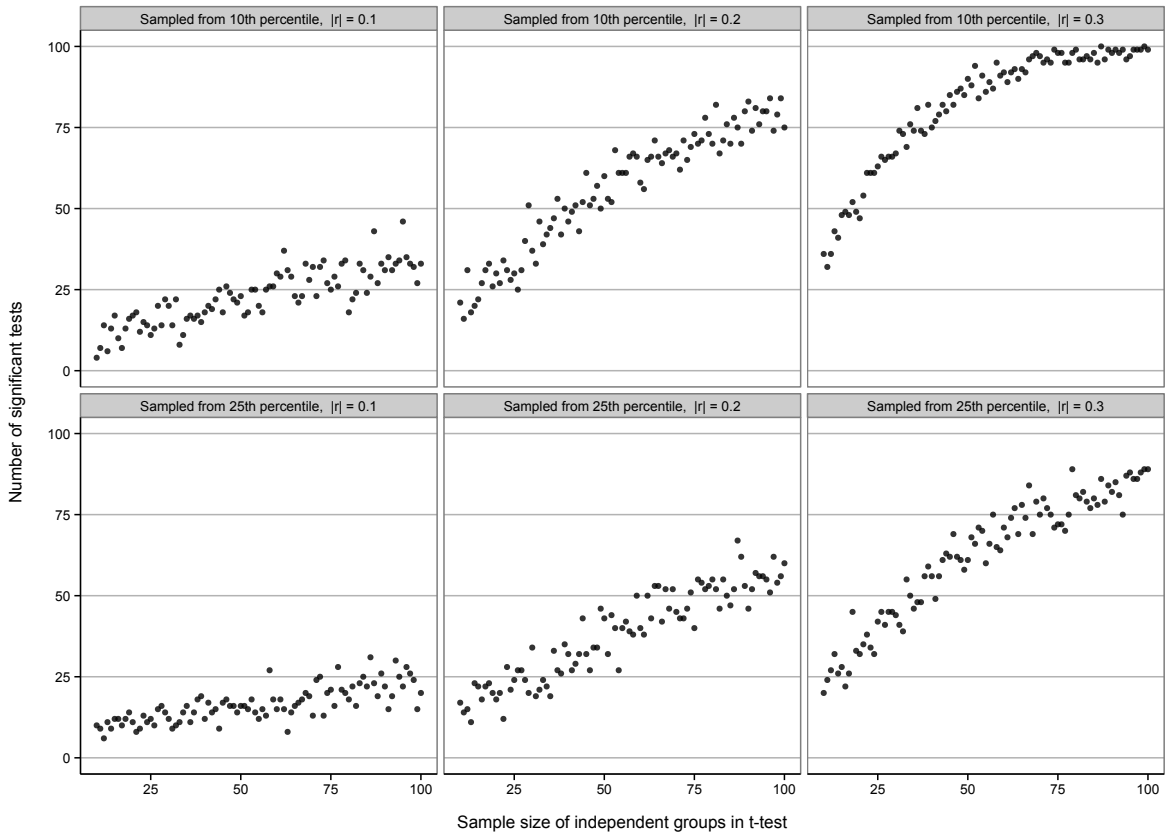

Figure 1.5. A simulation of the effect of sampling from different regions of a population distribution (N_{pop} = 500000) in the presence of a crud factor, a population-level correlation between any two random variables. Each dot represents the number of significant results ($p < .05$) observed in 100 t-tests for independent groups of the size represented on the x-axis (10 – 100). Two random variables were generated for each population correlation: .1, .2, .3 (columns). One random variable was used to sample data points in the 10th (top row) or 25th (bottom row), or between the 25th and 75th percentile (comparison group). The means concern the aggregated values of the second random variable for each sampled case. The directional hypothesis tested against the null was $(M_{[.25,.75]} - M_{[0,.10]}) > 0$ or $(M_{[.25,.75]} - M_{[0,.25]}) > 0$.

conduct a *t* test for the difference between dyslexic and average readers in the sample on their mean value of y (e.g., amount of music training received). The sample size for each group was varied from 10 to 100 data points and 100 tests were performed for each group size. For each test a new random group sample was from the different regions of the population distribution.

The graphs represent the number of significant ($p < .05$) *t* tests found in the series of 100 tests conducted for each group size. If the correlation between any variable were .1, comparing to the samples from the 10th and 25th percentile would yield 25% significant results at group sizes of 44 and 58 data points, respectively. The total study sample size would be 88 and 116. At this crud factor level the chances do not get much better than 1 in 4 corroborative events without there being any theory to pat on the back and grant some verisimilitude. When the correlation is .2, 25% significant tests can be expected at group sizes of 12 (10th) and 23 (25th) and at a correlation of .3 it's 10 (10th) and 12 (25th) participants in each group to find 25% significant differences. The crud factor of .3 even implies that 100% of the conducted tests could give a significant result if group size is larger

than 87 and the dyslexic group is drawn from the 10th percentile of the population distribution of reading ability.

The neglected credibility hurdle for evidencing empirical phenomena raises an important point: Predictions of phenomena have to be risky (Popper, 1959), ideally they should be point-values, magnitudes, but interval predictions should also be possible (Cumming, 2012; Meehl, 1997). Predictions have to be severely tested (Mayo & Spanos, 2006), but perhaps the best characterisation that is relatively school-of-philosophy and discipline-of-science free, was given by Salmon (1984): A scientist has to make sure it would have to be a "damn strange coincidence" to have observed the predicted phenomenon, *without* guidance by the theory. Perhaps we should just acknowledge that in the presence of the crud factor, even if it is just .1, evidencing a significant difference between average and dyslexic readers on whichever performance measure you can imagine, is not that impressive. It does *not* qualify as a "damn strange coincidence" and therefore not as a corroborator of theories: Enter my characterisation of the state of the current scientific record on causes of developmental dyslexia.

1.3.2 Investigator James corroborates his muse: The EVA-ætiology

The crud factor is one of the many important factors listed in Table 1.3 that make the appraisal of theories in the context of weak predictions and NHST problematic, if not impossible. One more general cause underlying this notion are the unknowns, such as the number of pilot studies that were conducted prior to the publication of the corroboration of a prediction (factor 7, Table 1.3). This is in principle the same effect as just demonstrated in Figure 1.5, except the decision where to place the cut-offs is purely guided by a previous sample of values that turned out to yield significance of the hypothesis test.

There are other reasons than statistical ones, to be explicit about assumptions, expectations, unknowns and every decision that leads to a change of method or procedure (i.e., the opposite of being obscurantist). It is essential for understanding the deductive chain that led to a prediction, its logical structure (if any) and which link should be to blame, or be rewarded when the results are known. As a demonstration, I sketch an example of the evaluation of a novel ætiology for developmental dyslexia. As far as I know, this ætiology has not been proposed yet, but all the claims and derivations are based on the actual scientific record. To prevent any confusion about my actual scientific claims (e.g., due to out of context quoting), I will present the evaluation as a description of the theory of Investigator James, $\mathbf{T_{ij}}$[13]

Investigator James published a paper in which a prediction of the EVA-ætiology of developmental dyslexia, or, *Early Vestibular-Auditory deficit hypothesis* is tested. Recently a "strong, early vestibular-auditory interaction that is critical for the development of music behaviour" was evidenced in 7-month-old infants, who were able to encode rhythm in music using their body as evidenced by their bouncing behaviour. Encoding only took place if they could actively bounce by themselves while listening to the music. The authors suggest that "the experience of body movement plays an important role in music rhythm perception" (Phillips-Silver & Trainor, 2005). According to Investigator James this allows the derivation of a causal pathway for understanding developmental dyslexia, because:

1) $\mathbf{T_1}$: Rhythm, beat and meter perception in non-speech, speech and music is impaired in dyslexic readers as well as rhythm production (Corriveau & Goswami, 2009;

[13] Investigator James is a zealous, but honest researcher who produces so many theories they have to be indexed T_{ij}. He is my terrestrial equivalent of Omniscient Jones, whose theories (T_{OJ}) are always true (Sellars, 1956).

Goswami, 2006, 2011; Huss et al., 2010; Overy, 2003; Ramus, Nespor, & Mehler, 2000; Thomson & Goswami, 2008) Results are accounted for by a theory proposing the real cause of dyslexia concerns timing deficits understood in a **neurobiologically** inspired **temporal sampling framework** (Goswami, 2011).

2) **T_2**: Balance and motor learning are impaired in dyslexic readers due to a **cerebellar deficit** (Brookes, Tinkler, Nicolson, & Fawcett, 2010; Ramus et al., 2003; Reynolds et al., 2003; Rochelle & Talcott, 2006; Stoodley et al., 2005; Stoodley & Stein, 2012). Exercise based training ameliorates language and literacy difficulties of dyslexic readers (Franceschini et al., 2013; Reynolds et al., 2003; Reynolds & Nicolson, 2007).

3) **T_3**: In average developing children and adults, a larger amount of lifetime music training is associated with better auditory discrimination, fine motor skills, a larger vocabulary, better non-verbal reasoning ability, better letter recognition and better speech in noise perception (Forgeard, Winner, Norton, & Schlaug, 2008; Proverbio, Manfredi, Zani, & Adorni, 2013; Strait, Parbery-Clark, O'Connell, & Kraus, 2013). A novel theoretical framework about **training-related plasticity** induced by the complex nature of learning, listening and playing **music** provides a **neuroscience explanation** for these associations (Herholz & Zatorre, 2012).

4) **T_4:** There is a **genetic factor** associated to impaired **speech perception and language production** (vocabulary) in infants of dyslexic parents, noticeable as early as 2-17 months of age (Been, van Leeuwen, & van Herten, 2008; Koster, Been, & Diepstra, 2005; Richardson, Leppänen, Leiwo, & Lyytinen, 2003; van Herten et al., 2008; van Leeuwen et al., 2006; van Leeuwen et al., 2008). These impairments in infancy are associated to reading and spelling problems later on in life (Guttorm, Leppänen, Hämäläinen, Eklund, & Lyytinen, 2011; Molfese, 2000). Speech perception based training ameliorates language and literacy difficulties of dyslexic readers (Tallal, 2004).

5) **T_5**: The genetic factor must be at work to impair auditory and speech perception in utero. It is known that there is **learning-induced plasticity of speech processing** before birth (Partanen et al., 2013) in addition to **speech perception**, **voice-pattern recognition** and **language learning** (DeCasper & Spence, 1986; Moon, Lagercrantz, & Kuhl, 2012). The cognitive and perceptual abilities that are higher in the population of well-trained musicians are typical of the abilities that are much lower in the dyslexic reader population. Given the conjectures of the musical training-related plasticity framework, it is likely that the temporal sampling theory and the cerebellar theory can be unified into a **general deficit in the fluency of coordination of perception and action**. Such a deficit was recently evidenced by examining long term correlations in naming latencies of dyslexic readers (Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012).

Using these 5 theoretical claims Investigator James believes he can predict an observational constraint between two unlikely variables that can finally resolve the long-standing debate on whether there is a speech perception deficit in developmental dyslexia (Ramus & Szenkovits, 2008; Serniclaes & Sprenger-Charolles, 2006):

> **Conjecture**: *From* the theoretical claim about importance of early vestibular-auditory interaction and the experience of self-initiated bodily coordination to musical rhythm as crucial factors for the development of music behaviour and rhythm perception *it follows that* impaired music and speech rhythm perception together with impaired balance and motor learning skills in dyslexic readers are also due to the reduced in

utero/neonatal/infancy learning-induced plasticity due to the general deficit of fluent coordination (**T_5**).

Prediction: *If* the amount of training-related plasticity induced by engaging in music related behaviour is associated with better letter recognition, a larger vocabulary and better auditory and speech perception in non-dyslexic readers *and if* action and speech training-induced plasticity can ameliorate reading ability of dyslexic readers *then* there must be a positive association between amount of music training received and speech perception performance in dyslexic readers (**$T_1 + T_2 + T_3 + T_4 + T_5 +$ conjecture**).

Most of the theoretical entities used by Investigator James touch the data base, the empirical record of the field, but he posits some novel functional or compositional laws to exist between entities, like **T_5** and the implication in the conjecture. The prediction is derived using at least five theoretical claims, each consisting of several auxiliary hypotheses and theories. Following Lakatos' terminology I will call the theoretical whole (**$T_1 + T_2 + T_3 + T_4 + T_5$**) the *hard-core* of the EVA-ætiology, or its *theoretical core* **T_{ij}**. Box 1.1 shows the elements involved in corroborating the theory. The sentence can be read as: "from **T_{ij}** *and* auxiliary theories *and* auxiliary instruments *and* assuming no interference of unknown factors and causes (*ceteris paribus* clause) *and* a truthful report of what transpired during the observation, *it follows that if* **O_1** is observed *then* **O_2** must also be observed." The horseshoe is the symbol for material implication and the turnstile represents a deduction of the material implication based on the left hand side.

O_1 concerns evidencing developmental dyslexia and **O_2** concerns observing a positive correlation between amount of music training received and speech perception performance. The four possible theoretical outcomes for the observations are that **O_1** and/or **O_2** can be evidenced or not at all. However, the consequences for **T_{ij}** concern just two outcomes, being whether the material implication is true or false. By applying the valid logical syllogisms (see Table 1.5), it is clear there can be only one valid logical inference when appraising a scientific theory: If the right hand side is false, then **T_{ij}** is false (*modus tollens*). The other valid syllogism (*modus ponens*) is the derived prediction: **T_{ij}** is true and therefore the right hand side is true (the observational constraint). This is Investigator James' *assumption*, his *theoretical claim*, it is the *hypothesis* he wants to test. Only Omniscient Jones' (T_{OJ}) theories are true (Sellars, 1956) and therefore all a scientific community can do is decide how truth-like **T_{ij}** appears to them which means assessing its verisimilitude.

Let **O_1** and **O_2** be observed, meaning the right hand side of the corroboration formula is true. The only valid conclusion is that **T_{ij}** is not falsified. To claim to have evidenced the truth of **T_{ij}** is to commit the invalid third figure *affirming the consequent*. This is probably the most often committed invalid inference made in science and known as the "effect = structure fallacy". It is often committed when inferring a cognitive modules exists based on each behavioural effect one observes (Bosman, Cox, Hasselman, & Wijnants, 2013; Van Orden & Kloos, 2003; Van Orden, Pennington, & Stone, 2001). Or, perhaps positing causal pathways for developmental dyslexia for each observed impaired performance measure evidenced in dyslexic readers as a group? Note that Investigator James, to derive the theoretical core of the EVA-ætiology, uses a mix of the syllogisms, mostly affirming the consequent, as if using modus ponens. This is based on a track record of just several corroborative events (the empirical studies cited), whose weight of corroboration we do not know much about. The ontological relevance of genetic and brain components does seep through in the weight attributed to some links in the derivative chain.

To be scientifically credible, or plausible, to enjoy a high perceived verisimilitude, the only thing $\mathbf{T_{ij}}$ can do is to make risky predictions that are severely tested. Not just once, it has to show a vita without any gaps, a track record of corroboration events (e.g., at least the 5 corroborative events at proper Total Power mentioned in the previous paragraph). More weight will be added to the corroboration if, *absent* $\mathbf{T_{ij}}$ the probability of observing $\mathbf{O_2}$ *conditional on* $\mathbf{O_1}$ is very low or is considered a "damn strange coincidence". Known factors such as the crud factor and weak directional hypotheses should greatly diminish the corroborative weight awarded to observing the predicted observational constraint. In the case of $\mathbf{T_{ij}}$ one should ask: How extraordinary is the observation of a positive association between music education and speech perception in dyslexic readers, without $\mathbf{T_{ij}}$ suggesting the empirical inquiry? Perhaps there is some surprise left after discounting for the crud factor and the weak prediction, it is for the community to discuss. What would have guaranteed high corroboration strength without any discussion at all would be Investigator James' prediction of the exact value of the positive correlation, conditional on the severity of dyslexia of a participant (a point-value prediction). Unfortunately, that is beyond the scope of current theory.

1.3.3 When is a Lakatosian defence defendable?

What happens if the right hand side is false (the case of the falsified conjunction in Box 1.1)? A consequence of the logical structure of the formula is that the falsification by modus tollens affects all the elements in the left hand conjunction that give rise to the derived prediction of observations. This means that the "cause" for not observing the predicted observations, the falsification, can apply to each (at least one) individual element in the formula. This can be called a "Lakatosian defence" as it represents the important amendments Lakatos added to Popper's logic of scientific discovery (Lakatos, 1970, 1974; Meehl, 1990b; Popper, 1959). Oversimplifying, in Popper's version of science, there would only be a **T** and modus tollens would dictate scientist to abandon the theory in an act of "instant rationality" (Meehl, 1990a). As in the case of Avogadro's number, molecular theory was not abandoned because the estimates weren't exactly the same. Some of the auxiliaries did not enjoy as much perceived verisimilitude as others and the community knew this. What eventually happened was that auxiliary theories used in the derivation chain were amended or replaced by more accurate versions which give the very accurate number presented earlier.

A Lakatosian defence allows a community of scientists to examine what in the derivation of the prediction might have caused the failed observation, without immediately abandoning the theory. It is possible a theory may need to be amended or that experimental procedures or instruments need to be refined. In my opinion this is a sensible thing to do in a science that has a tradition of seeking rigorous tests of theories that would maximise the corroborative strength of an observation, like a program of *strong inference* (Platt, 1964). Crucial experiments devised to test divergent predictions of competing theoretical claims. What is needed is a science that is able to keep track of the verisimilitude of the theories it produces in a more or less formal way (use meta-theoretical tools). Or, it has to be relatively clear to all practitioners of the science in question whether a theoretical prediction is supported by the data or not and which finite set of testable factors could cast doubts on that assertion (i.e., physics). This ensures that eventually the options for deflecting a modus tollens falsification death ray of instant realism to other factors besides T will be depleted in a relatively short period of time. In a science that does not formally and rigorously appraise its theories a situation could arise that should be somewhat familiar to the reader by now: Many different theories, apparently corroborated by empirical evidence, but the evidence

The corroboration formula for appraising a theory:

$$(T \cdot A_T \cdot C_p \cdot A_I \cdot C_n) \vdash (O_1 \supset O_2)$$

T : The theory of interest
A_T : Auxiliary theories relied on in the particular experiment
C_p : Ceteris paribus clause (other things being equal)
A_I : Instrumental auxiliaries (devices relied on for control and observation)
C_n : Realised particulars (conditions were as the experimenter reported)
$O_{1,2}$: Observations, or statistical summaries of observations

The case of the falsified conjunction $(O_1 \cdot \sim O_2)$:

$$\sim (T \cdot A_T \cdot C_p \cdot A_I \cdot C_n) \equiv \sim T \vee \sim A_T \vee \sim C_p \vee \sim A_I \vee \sim C_n$$

Box 1.1. The corroboration formula for appraising a theory, adapted from (Meehl, 1990a; 1990b; 1997; 2002). The outcome in the case of a falsified conjunction (observation 2 was not evidenced) is given as two forms, one falsifying the conjunctions and its logical equivalent, falsifying at least one of its elements. Explanation of symbols: The dots state conjunctions ("*and*"), the ∨ disjunction ("*or*"), the turnstile (rotated T) signifies deductive derivability (entailment, "*from … it follows that …*"), the horseshoe (⊃) is a material implication or conditional ("*If … then …*"), a tilde (~) represent negation ("*not*"), the three horizontal bars indicate full equivalence of the left and right hand side.

has no authority to discorroborate or negatively affect the perceived truth-likeness of other theoretical claims.

Returning to Investigator James' theory, let the observational conjunction be falsified, the positive correlation was not observed in dyslexic readers. What are the credible options for saving $\mathbf{T_{ij}}$ from refutation? What is often observed in empirical reports of the social sciences is that causes for a (partial) failure to corroborate a prediction are attributable to unknown or unexpected factors that may have influenced a result. This means the *ceteris paribus clause* ($\mathbf{C_p}$) is sacrificed in order to save **T**. $\mathbf{C_p}$ states that under the conditions of the test (e.g., the experiment), there are no causes other than those in the derivation chain that could lead to the predicted observation. Assuming $\mathbf{C_p}$ is *true* prior to the experiment is extremely important for the corroboration of the derivational chain. Assuming $\mathbf{C_p}$ must have been *false* after a failed test is the easiest way out of falsification of **T**. Common statement are: "sample was too heterogeneous", "effect was observed after including an [ad hoc] moderator". In the field of developmental dyslexia one can find discussion sections in which authors ask whether the dyslexics were "real" dyslexics,

whether IQ was taken into account, biological age vs. academic or reading age, comorbidity. Chapter 2 and subsequent chapters provide a possible explanation for the easy and apparently credible $\mathbf{C_p}$ sacrifices and suggest context relativity of measurements should be a part of the derivative chain of observations. It has been known for some time that the tendency to make sure that "all other things being equal" is true, by imposing strict procedures for sampling, selecting and matching of participants, and so forth, basically means that verisimilitude will apply to that context alone. For studying properties of particles in a collider, that may be fine, but when a medical procedure has to be introduced into the public domain, it often turns out that theoretical causes (a pharmaceutical substance) proven very effective in a randomised control trial (RCT), turns out to be much less effective (= efficient) when introduced into society (Cochrane, 1972; Higgins, Green, & Collaboration, 2008).

The $\mathbf{C_n}$ part of the equation has received a lot of recent attention in psychological science, some of which were be briefly discussed in the preface (e.g., questionable research practices, *p*-hacking). The essential point is, if one cannot trust authors to provide full disclosure about everything one needs to know about the particulars of the predicted observation, its verisimilitude cannot be assessed (a recent call for public disclosure of particulars revealed some authors may violate this clause without even knowing it, see Lebel et al., 2013).

For $\mathbf{T_{ij}}$ and the EVA-ætiology the degree of truth-likeness of the theories that have been used to derive the prediction are important. These auxiliary theories and hypotheses ($\mathbf{A_T}$) have been used to infer 'modus ponens'-like relations connecting different theoretical entities (e.g., functional or compositional laws), without actually appraising them. In addition, the verisimilitude of the entities is unlikely to be known in any formal way, as in a track record of known precision and accuracy of corroborative events. In the case of $\mathbf{T_{ij}}$ the theoretical core consists of five theoretical statements (i = 1, … 5) and the conjecture "binds" them in a 'modus ponens'-like manner in order to allow the actual prediction of the observational conjunction. Within each of the five postulates that constitute the core, there are many auxiliary hypotheses and theories (indexed by j = 1, … n). I estimate on average 5 theoretical claims or events that corroborate a hypothesis within each core postulate are important enough to be blamed for not observing what was predicted. Ranging from the inference of a genetic factor based on infant prospective studies of children at risk for dyslexia to the increased speech in noise perception in non-dyslexic readers due to music training. There are at least 25 possible targets to which refutation may be deflected before Investigator James' actual prediction or the theoretical core is affected. Lakatos called this the protective belt, preventing the core from falsification. If ad hockery is allowed in a science, the protective belt of sacrificial auxiliary hypotheses is of course infinite. In the spirit of this book, I will use *protective boundary* to denote this phenomenon and proclaim that tests of predictions and appraisal of theories, should aim *beyond the protective boundary,* straight for the core. A sobering note about this objective, technically, the reports of corroborative events that constitute a substantial part of the 25 possible auxiliary sacrifices, will have a similar corroboration formula associated with the observation, the auxiliary takes on the role of **T** but with its own personal stock of sacrificial of auxiliaries. It almost seems as if there were a dark plot to keep us all occupied indefinitely.

Finally, one might suggest the use of meta-analysis as a tool to gain some confidence about the credibility of an entity or law. These analyses report a summary effect size of a particular predicted phenomenon, and magnitudes of effect sizes are generally not predicted by theories, mostly signs of associations. Using a meta-analysis for these purposes

Table 1.6

The Four Figures of Deductive Inference. Two Syllogisms Constitute Valid Logical Inferences.

Figure	Statement	Name	Conclusion
1	`if p then q, p ∴ q`	Modus ponens ("establishing mode")	*Valid*
2	`if p then q, ~p ∴ ~q`	Denying the antecedent	Invalid
3	`if p then q, q ∴ p`	Affirming the consequent ("effect = structure fallacy")	Invalid
4	`if p then q, ~q ∴ ~p`	Modus tollens ("destroying mode")	*Valid*

(to gain credibility for an effect) is committing the effect = structure fallacy quite literally, *because* no effect magnitude has been predicted except for a lower bound, the critical value associated with the α level. The summary effect sizes can be interesting to asses after the fact, to see if a study included in the meta-analysis was sensitive enough to have detected the summary effect size (Button et al., 2013). As a corroboration of predictions meta-analysis does not yield any empirical accuracy or precision outcomes that should comprise the track record of the theory. From the meta-theoretical perspective, meta-analysis can help to improve precision and sensitivity of future measurements by quantifying the sources of variation between studies that are supposed to measure the same phenomenon.

Meanwhile, Investigator James' eagerly awaits the reviews of his new and improved funding proposal to further test the EVA-ætiology, using RCTs this time. A systematic review of 851 studies on the relation between music education and language and literacy development of dyslexic readers had of course nothing to do with his design change and new focus of inquiry:

> "There is no evidence available from randomized controlled trials on which to base a judgment about the effectiveness of music education for the improvement of reading skills in children and adolescents with dyslexia. This uncertainty warrants further research via randomized controlled trials, involving a interdisciplinary team: musicians, hearing and speech therapists, psychologists, and physicians." (Cogo-Moreira et al., 2012)

1.4 The tenuous nomological net: Theories of construction vs. theories of principles

The problems with theory evaluation presented so far are not new and can be appended to a long list of critiques tracing back to the earliest conceptions of some fields of scientific inquiry. For example, Ladd (1892) reviewing William James' "The Principles of Psychology" (1890) concluded that establishing psychology as a natural science was an "utter impossibility". James' had suggested that psychology was already using the methods of the natural sciences to test deep hypotheses about its object of study, while in according to Ladd:

> "[…] psychology as a science, *devoid of all postulating of* "*deeper-lying entities,*" does nothing of the kind. It *assumes only the phenomena* - the thoughts and feelings as actually known, and *the possibility of ascertaining uniform relations* among them." (Ladd, 1892, pp. 29–30, emphasis added)

Based on some of the work discussed in the previous paragraphs, Lakatos classified the kind of theorising practiced in the empirical social sciences as one of the worst kinds of *ad hockery*[14] and predicted grave consequences if it were allowed to continue:

> "After reading Meehl [1967] and Lykken [1968] one wonders whether the function of statistical techniques in the social sciences is not primarily to provide a machinery for producing phoney corroborations and thereby a semblance of 'scientific progress' where in fact, there is nothing but an increase in pseudo-intellectual garbage. [...] Or, as Lykken put it: 'Statistical significance [in psychology] is perhaps the least important attribute of a good experiment; *it is never a sufficient condition for claiming that a theory* has *been usefully corroborated*, that a meaningful empirical fact has been established, or that an experimental report ought to be published.' [...] Thus the methodology of research programmes might help us in *devising laws for stemming this intellectual pollution which may destroy our cultural environment even earlier than industrial and traffic pollution destroys our physical environment.*" (Lakatos, 1975, p. 176, footnote 1, emphasis added)

Why do the softer fields of science rely on NHST, or more generally speaking, significance testing of directional predictions to "ascertain facts", when it has been pointed out (by the very people who invented the techniques!) that to do so is logically flawed, mostly trivial and often just plain wrong? (Bakan, 1966; Carver, 1993; Cohen, 1994; Hogben, 1956; Lykken, 1968; Mayo & Spanos, 2006; Meehl, 1967; Michell, 2009; Neyman, 1969; Nix & Barnette, 1998; Nunnally, 1960; Ring, 1967; Rosnow & Rosenthal, 1989; Rozeboom, 1960; Steiger, 2004; Trafimow, 2003; Tukey, 1960a, 1960b; Wilkinson, 1999). Why doesn't it look for better formal tools, or just hire a mathematician to create them? In fact, it was sir R.A. Fisher himself who explicitly warned about the dangers of adopting the abstract mathematical concepts of the theory of probability and measurement error, without carefully examining whether they are appropriate for social science. That is, in service of the goals of improving "natural knowledge" about relevant phenomena, as is customary in the natural sciences:

> "I am quite sure it is only personal contact with the business of the improvement of natural knowledge in the natural sciences that is capable to keep straight the thought of mathematically-minded people who have to grope their way through the complex entanglements of error, […] Certainly there is grave confusion of thought. We are quite in danger of sending highly trained and highly intelligent young men out into the world with tables of erroneous numbers under their arms, and with a dense fog in the place where their brains ought to be." (Fisher, 1958, p. 274; also see Yates, 1968, who reiterates Fisher's point a decade later).

In the sections that follow I will attempt to explicate Fisher's "grave confusion of thought" about "the improvement of natural knowledge" in the soft sciences. These observations warrant a broader discussion of problems with theory construction in the soft sciences before suggesting novel directions for the scientific inquiry into the ætiology of developmental dyslexia. We need to respect our elders (and sometimes their frustration and discontent and perhaps their failures as well), because history *is* repeating itself and this time there is a tendency to feign ignorance about the severity of non-replicability of phenomena, hypothesising after the results are known (HARKing; Kerr, 1998), underpowered studies, *p*-hacking and data-peeking. I am not citing opinions by famous scholars to taunt the scientist of the softer fields.[15] I cite them to point out that claiming

---

[14] There is also honest ad hockery: "In our theories, we rightly search for unification, but real life is both complicated and short, and we make no mockery of honest ad hockery" (Good, 1965)

[15] This appears to be a sentiment among some peers: "replication bullies" and "data detectives" ruin their careers by pointing out statistical errors in their published work. (See e.g., Schönbrodt, 'About Replication Bullies and Scientific Progress… Retrieved May 2014 from: http://www.nicebread.de/about-replication-bullies-and-scientific-progress/)

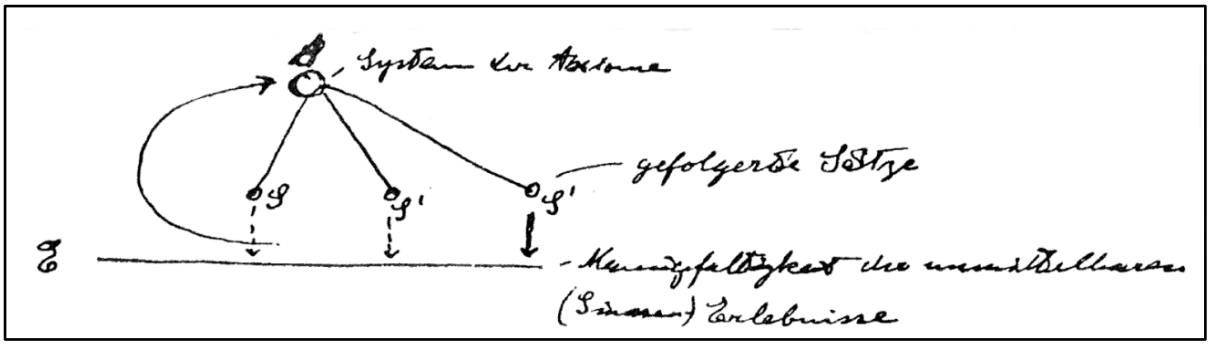

Figure 1.6a. A drawing by Albert Einstein explaining his philosophy of science. Taken from a letter sent to Maurice Solovine, 7th of March 1954 (cf. van Dongen, 2010).

ignorance on these issues in present, is to blatantly deny a historical scientific record, laden with warnings by statisticians, philosophers, theorists and methodologists about making invalid inferences based on empirical data. The fact is there still isn't a datum universally agreed on (Spencer, 1855) and although encouraging practices like data sharing or conducting confirmatory and replication studies is very necessary, it will not suffice to save the empirical social sciences from itself. One must ask, what is so special about the way the natural sciences theorise about the constituents of reality?

1.4.1 Constructive and Principle Theories: Synthesis versus Analysis

At the time of the 'confirmation' of some of the testable predictions of the theory of General Relativity discussed in paragraph 1.3, Einstein wrote a lucid letter to the London Times (November 28, 1919) in which he characterised two kinds of scientific theory:

> "We can distinguish various kinds of theories in physics. Most of them are constructive. They attempt to build up a picture of the more complex phenomena out of the materials of a relatively simple formal scheme from which they start out. Thus the kinetic theory of gases seeks to reduce mechanical, thermal, and diffusional processes to movements of molecules – i.e., to build them up out of the hypothesis of molecular motion. *When we say that we have succeeded in understanding a group of natural processes, we invariably mean that a constructive theory has been found which covers the processes in question.*
>
> Along with this most important class of theories there exists a second, which I will call "principle-theories." These employ the analytic, not the synthetic, method. The elements which form their basis and starting-point are not hypothetically constructed but empirically discovered ones, general characteristics of natural processes, principles that give rise to mathematically formulated criteria which the separate processes or the theoretical representations of them have to satisfy. *Thus the science of thermodynamics seeks by analytical means to deduce necessary conditions, which separate events have to satisfy, from the universally experienced fact that perpetual motion is impossible.*
>
> The advantages of the constructive theory are *completeness*, *adaptability*, and *clearness*, those of the principle theory are *logical perfection* and *security of the foundations*."
>
> – Einstein (1934\1952; emphasis added)

Contemporary physical theories are mostly principle theories and although he described constructive theories as "most important" it is evident for example from the contents of a letter to a friend (see Figure 1.6a) that Einstein believed only principle theories could advance fundamental scientific knowledge about the universe.[16] To explain the profound difference between the two types of scientific theory, the metaphor of the

[16] I thank Michael Seevinck for introducing me to this important history of science.

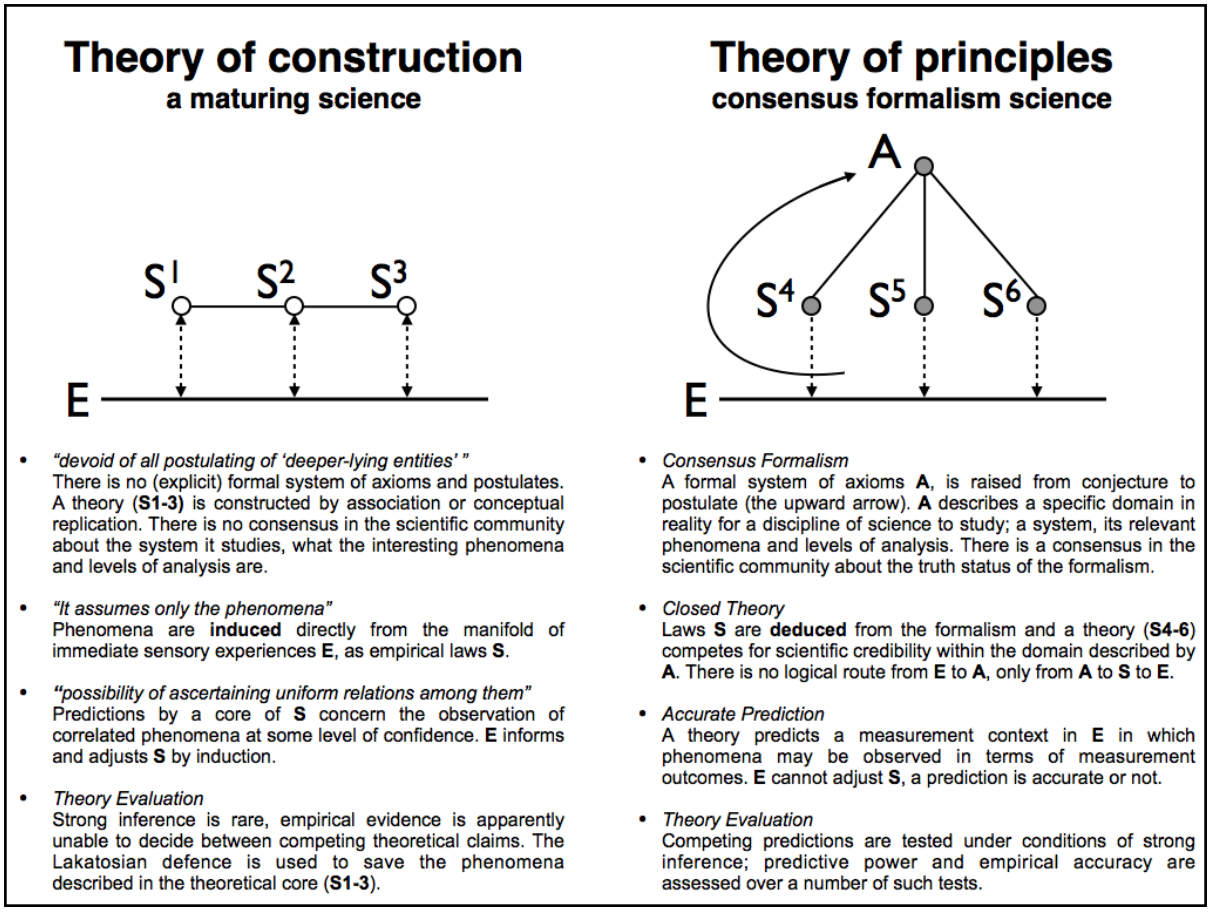

Figure 1.6b. An adaptation of the "Solovine Schema" shown in Figure 1.6a (see van Dongen 2010, p. 52-53). The quoted text is from Ladd's 1892 characterisation of James' vision of psychology as a natural science.

nomological net is again helpful. A part of the nomological net representing a scientific description of a domain in reality that is logically strong with a track record of strong corroboration and thus perceived as having high verisimilitude, may look something like the right panel of Figure 1.6b. This would be a rather coarse rendition of a nomological net as the nodes **A** ("System der Axiome" [system of axioms]) and **S** ("Gefolgerte Sätze" [deduced laws]) represent collections of many different kinds of theoretical entities. As explained earlier, in terms of the nomological net metaphor (cf., Cronbach & Meehl, 1955), the strands of the net are supposed to represent the functional or constitutive laws that connect theoretical entities. Turning to a more stylised version in Figure 1.6b (right column), it is important to note that from **A**, lawful relations between theoretical entities **S** can be deduced and those lawful relations predict observable phenomena in **E** ("Mannigfaltigkeit der unmittelbaren (Sinnes-) Erlebnisse" [manifold of immediate / direct sensory experiences]). One or more **S** together can be regarded as a theoretical whole, a scientific theory (e.g., $\mathbf{S}_{4\text{-}6}$), the theoretical core.

What makes these theories principled theories? Perhaps surprisingly, it is not the logical strength of a pathway from observations of phenomena in **E** to a system of axioms **A** formally describing a domain in **E**. According to Einstein such a logical route from **E** to **A** does not exist at all. The process of theory construction always starts with a creative-intuitive act in which a scientist raises a conjecture to a postulate (cf., van Dongen, 2010). This is the

upward arrow from **E** to **A** (the system of axioms)[17] and it represents the bold, but inspired act of a scientist laying explanatory claim to a certain domain in reality, by stating: "let's see what happens if we assume it really is like this". As mentioned earlier, in the hard sciences it is often the case the definition of a formalism is a community effort and therefore there is consensus about its verisimilitude. The purpose of a scientific theory departing from within the domain defined by the axioms and postulates of the formalism is to predict observable phenomena in **E** with a higher precision and accuracy than predictions by competing theories that depart from the very same formalism.

The actual test of the precision and accuracy of the predicted phenomena (e.g., by running an experiment, conducting observations) constitutes an appraisal of the verisimilitude of the theory. Its outcome should ultimately have consequences for the perceived truth-likeness of the entire formalism as well. Theories based on the same principles allow for a program of strong inference (e.g., Platt, 1964): A single datum can be obtained in a crucial experiment to decide between the veracity of competing theoretical claims. The logical strength of the derivational chain leading to the prediction is such that "It is the theory that decides what may be observed" (Einstein as quoted by Heisenberg, 1971, pp. 62–63). The single headed arrows in the right panel represent explanation by D-N (Table 1.2) and as such, principled theories can be characterised as 'closed' (cf. Bokulich, 2004). They break down as a whole when a crucial experiment does not yield the results predicted by the theory: "[…] if an experiment does not fit in Newtonian physics, you don't know what you mean by the words." (Heisenberg interviewed by Kuhn, 1963, p. 24 , February 27th). In his letter to the London Times, Einstein makes a similar remark about GRT:

> "The chief attraction of the theory lies in its logical completeness. If a single one of the conclusions drawn from it proves wrong, it must be given up; to modify it without destroying the whole structure seems to be impossible." (Einstein, 1934\1952)

Of course, as discussed in the previous paragraph, for theories of such logical strength that can be shown to be false so conclusively, a Lakatosian defence is permitted if the circumstances warrant it.

In theories of construction, depicted in the left panel of Figure 1.6b, a theoretical whole (e.g., $\mathbf{S_{1\text{-}3}}$) is "hypothetically constructed" instead of "empirically discovered" by proposing associations between phenomena described in the empirical record exist. In the most exemplary case there are no restrictions whatsoever on forming associations between observed phenomena (Einstein called this "adaptability" in the quote form the letter). The left panel of Figure 1.6b represents a tenuous nomological net of theoretical entities that are all connected to, and therefore defined by empirical observations (exclusively I-S or sometimes D-S explanations, hence the double-headed arrows). There is no (explicit) abstract formalism defining a domain in reality that could prevent theoretical entities to become associated with one another in order to constitute a theoretical whole. This is why it cannot be said they are empirically discovered by testing risky predictions, their adaptability can make them "complete": Any theoretical entity can become associated to any other entity by constructive hypotheses. To identify a theoretical core, one will often look for the least disputed empirical results and the hypotheses that predicted them. The options to avoid refutation, as explained in section 1.3, will be virtually limitless if the epistemic link is severed. The protective boundary will specifically serve to save those phenomena that support the theoretical core.

[17] I will refer to the system of axioms as *the formalism* hereafter. A full mathematical axiomisation of a physical theory is rare (Bunge, 1967).

This bears strong resemblance to the situation described for theoretical accounts of the ætiology of developmental dyslexia in the previous paragraphs: The node categorising theory words was associated to all the other nodes that mainly captured different experimental designs or performance measures, hence empirical phenomena (e.g., brain, attention, learning, auditory perception). As a consequence, all the theoretical entities and laws that constitute the theory are very direct descriptions of empirical phenomena observable as a manifold of direct (sensory) experiences (e.g., "it assumes only the phenomena"). That is, in most cases, to describe theoretical entities in a theory of construction one can suffice with a level of abstraction that does not rise above a common language description: "speech-sound perception", "rules for converting mental representations of letters into sounds", "a part of the visual system capable of fast signal transmissions". Of course, in the literature these entities be referred to by their own proper neologism, but this is very different from the level of abstraction necessary to describe observables corresponding to non-commutative operators in a quantum measurement or a Ricci tensor in relativistic cosmology (both are mathematical abstractions that are hard to relate to any perceptual experiences). The derivational chain leading to a novel prediction in a theory of construction is weak, or soft and so are the topologies of its nomological nets, even so, "he who casts, will catch". In most cases the chain of 'derivations' is based on previously observed particular facts (e.g., "[…] ascertaining uniform relations among [phenomena]") associated by probabilistic laws (see Table 1.2). As a consequence, the failure to observe a predicted phenomenon does not nearly have as many implications for the perceived veracity of the theoretical entities involved, as is the case in a theory of principles. The scientist who uses theories of construction to study the structure of reality refrains from "postulating deeper-lying entities" (Fanelli, 2010) and constructs and amends a theoretical whole guided by phenomena directly observable in **E**. In a play of words with Einstein's dictum, in the softer sciences: "*It's the observations that decide what may be regarded as the theory*".

1.5 The thin ontic line (and why it must be cut)

1.5.1 This is the story so far:

In section 1.2 the impression that a plethora of different theoretical claims about the causes of developmental dyslexia are considered to be "true" at the same point in time or at least the impression that they escape falsification by means of discorroborative empirical evidence, was made plausible by a quasi-historical, quasi-nomological network analysis of terms used in abstracts of scientific articles on the subject.

In section 1.3 meta-theoretical causes were proposed to explain the erosion of the corroborative authority of empirical evidence. Important causes for the field of dyslexia research were identified such as the crud factor, epistemic sloughing by hypothesis testing and the inappropriate use of the Lakatosian defence.

Section 1.4 provided a meta-theoretical taxonomy of theorising in science and identified the theories produced by the social and life sciences as theories of *construction* that fail to achieve a level of abstraction of reality beyond natural language descriptions. They refrain from positing phenomena at deeper-lying levels of reality.

Which identified cause is sustained by which identified effect, or vice versa, is difficult to tell and most likely futile to attempt to figure out. One half of the arguments borrow authority from history, the other half from philosophy and logic, and sprinkled on

top are some extra crunchy authorisations from physics. But what does this strange concoction authorise? To me, it authorises the rest of this book, including its slightly ridiculous late coming of age. It authorises a critical and more formal examination of scientific theorising when applied to inquire about the developmental origins of an observable state of impaired performance in young children and of the technology developed to alter that state. First, I sketch the framework for theorising and theory appraisal in the softer fields of science that emerges from the story so far.

1) $\mathbf{T_{ij}}$ is posited by a scientist in the softer fields of science as a set of connected statements about entities and laws. The text in which $\mathbf{T_{ij}}$ is embedded is not an *operational text*. That is, there is no separate formalism containing theory language defining the entities and laws that an operational text commonly connects to observational language. There is no formal calculus or "truth-grinding machine" that can be used to decide on logical structure of derivations and verisimilitude of corroborations. The embedding text may be something akin to an interpretative text. This is a text in which theoretical concepts are defined in terms of the theory language alone. Theories of construction contain mostly observation language, so their embedding texts, that is, the scientific publications in peer-reviewed journals of soft empirical science, define operational theoretical entities in terms of other operational theoretical entities.

2) A meta-theorist can still extract a lot of information from the set of statements. A soft embedding text may contain some hidden theory language and a large subset of laws and entities will be associated to empirical evidence by observational predicates and statements. This often means formal language is introduced in corroboration (measures of association, functional forms, etc.).

3) $\mathbf{T_{ij}}$ contains postulates that can be categorised as central (i) or peripheral (j). There are just a few central, or *core postulates* and associated to each of them are many peripheral, or *auxiliary postulates* (A_T) that have to have some degree of truth-likeness. *Auxiliary instrumentation* (A_i) the theory relies on is also a part of the periphery. In the case of corroboration by prediction of empirical observations, a ceteris paribus clause (C_p) has to be adopted together with an assumption about the truth of statements about realised particulars (C_n). The truth status of the latter clauses represents the basic conditions for an empirical science to work. Improving assertions about their truth is the topic of many contemporary discussions in social science.

4) $\mathbf{T_{ij}}$ uses *entities* in derivational chains: Mental representations, genetic and neural components (structures, complex), developmental milestones (events), persistent disabilities (states), and realisable dispositions such as training induced plasticity to ameliorate language and literacy. Entities may be classified by a meta-ontology that I will refer to as mentology.[18] An example is "Meehl's ontology" (Meehl, 1993). According to this mentological list that has been useful for appraising psychological theories, the world consists of *substances*, *structures*, *events*, *states*, *dispositions* and *fields*. In softer fields, structures are almost always complexes, never simplexes like quarks. As different parts of a complex can have different verisimilitude, appraisal of

[18] Mentology could represent a contraction of Meehl's Ontology or Meta / Mental Ontology, but more importantly it represents the fact stressed by Meehl that any attempt at a general classification of the lynchpins of reality depends on the mental lynchpins of the scientist. Meehl frequently solicited suggestions for change, expansion or reduction.

such theoretical entities is difficult without formal definitions of the components of a complex.

5) **T_{ij}** uses *laws* to connect entities in derivational chains: A mental representation of speech sounds includes frequency and amplitude information (*compositional / structural law*, defining the parts and their arrangement). A mental representation that is not composed according to a law can cause a state of impaired reading ability (*functional / dynamic law*, change processes). In mentology there are also *developmental laws*, they represent a combination of the compositional and dynamic laws often pertaining to variables that are not under control, such as in the "documentary disciplines": history, palaeontology, geology and evolution. Darwin's theory contains a developmental law.

6) Laws can be statistical or formal in nature (see Table 1.2) and can associate entities by specifying: *the sign* of *1st and 2nd order derivatives*, *partial derivatives* (order of signs, factor interaction), *functional form* (exponential growth function, predator-prey dynamics), *rank order of parameter values*, or *actual parameter values* (point-value prediction). More on theory specification in 1.6.4.

7) **T_{ij}** may be represented as a nomological net. Entities constitute the nodes, the laws the edges. A proper subset of the net makes contact with reality in that it is an observational subset. Some entities can be directly related to observations by I-S inference without violating Einstein's dictum, or the logicians' corroboration formula. To logically re-*construct* theoretical propositions such that they acquire empirical content can be done using the so-called Ramsey Sentence. Einstein posited the creative intuitive act to cut loose from reality into abstraction, but 'something' empirical is always retained. This implicit phenomenon was named "Ramsified upward seepage" (Meehl, 1990; Meehl, 1978, 2002, 2004b). What is important to note is that according to meta-theory it is possible to *assert* and *define* an entity at the same time, not in the operationalist sense, but by using a system of formal expressions.

8) Falsification of a theoretical core in soft science does not appear to occur very often. The same holds for corroboration or discorroboration by replication of observations, because such studies are rarely conducted until very recently (Klein et al., 2014). Much of the perceived verisimilitude of core theories in soft science is based on corroborative singularities. Black Hole phenomena in the nomological net that can only be approached as far as the event horizon. Their structure remains obscured to the inquisitive scientist, no matter how much daylight is let in to illuminate it. More formally stated, a Lakatosian defence is used to deflect falsification of the core to auxiliaries, the periphery. Or, the truth status of the ceteris paribus clause is sacrificed and an ad hoc auxiliary hypothesis is postulated. If an identified C_p falsifier can be changed into a corroborator on subsequent inquiries, this is a progressive research program, because a new structure was uncovered. If the falsifier is only used to protect the theoretical core of the postulates and the Black Hole status of its singular corroborative events, this is a degenerative research program.

Figure 1.7 displays what the results might be if a more formal system of theory evaluation was adopted in the soft sciences, one that allows positing theories of construction that can be evaluated in progressive research programs and eventually can make point-value predictions about parameters as a theory of principles. In terms of

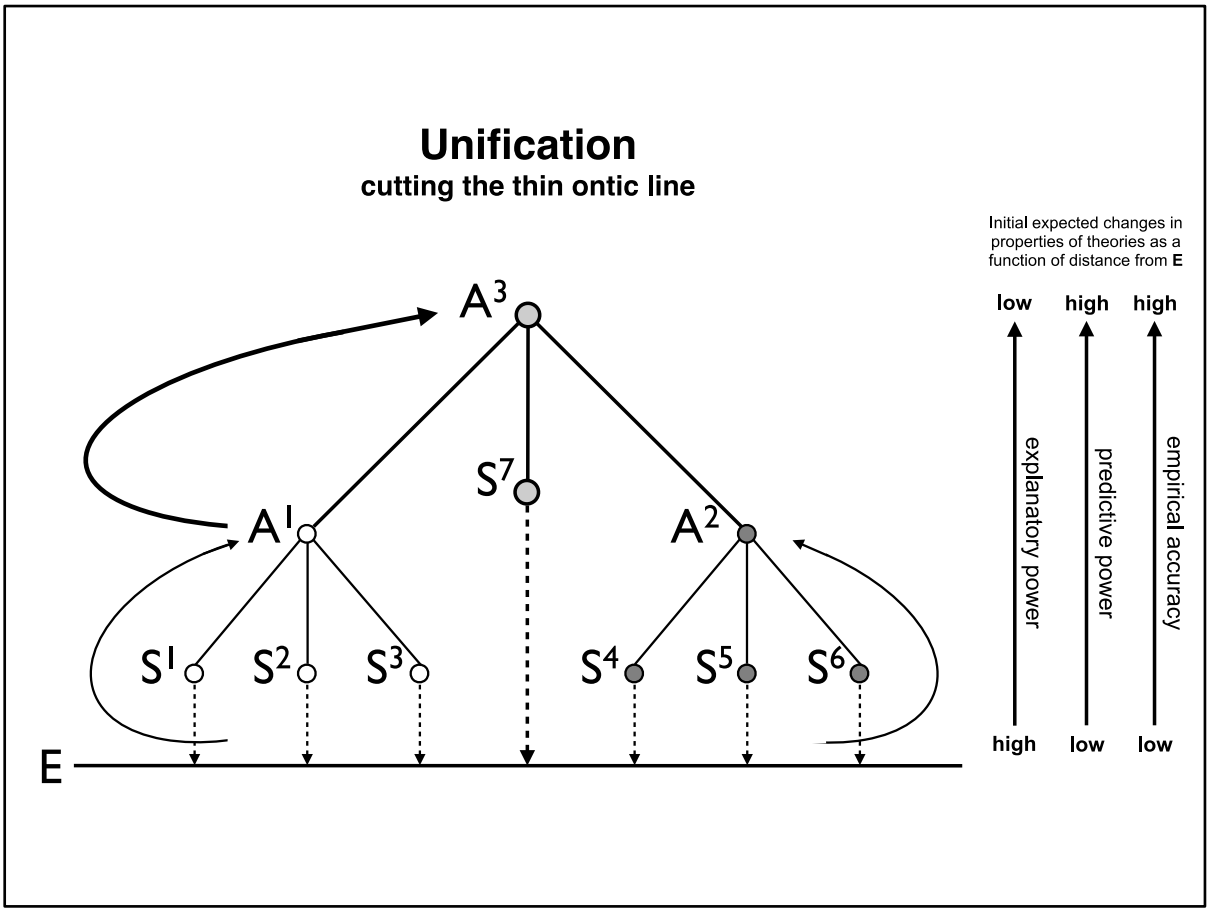

Figure 1.7 Einstein's unification (see van Dongen, 2010) and the consequences in terms of explanatory power, predictive power and empirical accuracy of theories by increasing distance from the manifold of direct experience, **E**.

corroboration, a track record of predictive power and empirical accuracy has to be established by a theory. The goal of scientific explanation should be unification, finding out how entities are the same, or can be understood by more fundamental, or sometimes more general laws, a progressive program. Einstein understood however that a unified theory "*pays for its higher logical unity by having elementary concepts […], which are no longer directly connected with complexes of sense experiences.*" (cf., Seevinck, 2011). To create a theory of principles means sacrificing a kind of explanation that is more a personal understanding of the world in terms of everyday experiences. This was not just a coincidental oddity for Einstein, he firmly believed that the fundamental notions of physics cannot be induced from experience and they cannot be justified a priori on the basis of our faculty of knowledge. This strongly opposes Kant's notion of incommensurability between Biology and Physics (e.g., Hasselman, Turvey, Seevinck & Cox, 2011; Seevinck, 2011).

There is a peculiarity about the role of ontology in this loss of explanatory power that I will leave for thorough discussion in the final chapter of this book in which I will also suggest how to (meta-theoretically) deal with its loss. It has to do with overcoming the enormous psychological satisfaction that is felt when empirical data agree with a cultural belief system or a personal world-view (Salmon, 1990). The danger is that an evaluation of the theory that predicted the data in a more or less formal chain of deductions may be positively biased towards such beliefs by its satisfactory agreement. This danger is quite literally quoted in the definition of the hard/soft science divide by Fanelli (2010): Socio-cultural and psychological factors decide how data are collected, analysed and interpreted. I

call this the *interpretation fallacy of theory evaluation* and it occurs very often in the social sciences, but is hardly ever noticed as problematic (Lebel & Peters, 2011 use "interpretation bias" in a slightly different, but related meaning).

1.5.2 Blinded by downward seepage? Cause and effect of the interpretation fallacy.

In the times before modern empirical science the existence-proof of ontological entities to our senses prompted ancient scholars to theorise about their role in the workings of the physical universe and the human body, like the classical elements and the Hippocratic humours. This is a very sensible and scientific thing to do if you have no highly corroborated nomological net to guide you. Theories were *constructed* based upon their empirical evidence as "measured" by our senses. This is essentially related to the logicians explaining implicit empirical content in theoretical propositions, "Ramsified upward seepage" (Meehl, 1978, 1990a; 2002, 2004b).

Here, we stumble upon the ancient problem of philosophy (mentioned in the preface): Is our internal world, the mind or our consciousness, part of reality? Should phenomena of the mind be allowed as a construct to theorise about the mind or should their existence to our mind's eye be ignored in theories? We must acknowledge that in social science, but psychological science in particular there is a thin ontic line between the theoretical concept of a mind and the actual sensory phenomenon of mind. Is the line too thin? For example, to the minds eye the external continuous universe appears filled with discrete objects, our internal thoughts, memories of events and the words we use to communicate are all discrete. Therefore, we may be naturally inclined to construct theories based on discrete components, without ever considering continuous architectures of mind as an alternative (Spivey, Anderson & Dale, 2008; Spivey & Dale, 2006). We experience time as a linear flow, memory, history, prospection, but what if it is not like that at all as some physical theories suggest (Brown & Uffink, 2001; Bunge, 1968; Nottale, 2010). Within the framework of ecological psychology, intentional dynamics and event perception are a very different conception of time and its passage is posited, a perspective known to very few soft scientists (Gibson, 1966; Kugler, Shaw, Vincente, & Kinsella-Shaw, 1990; Shaw & Kinsella-Shaw, 1988; Shaw & Turvey, 1999; Shaw, Flascher, & Mace, 1996; Turvey & Carello, 2012). Perhaps not surprising that the "inventors", like roboticists and industrial designers have a different opinion. In the final activity report of the research project MACS (Multi-sensory Autonomous Cognitive Systems Interacting with Dynamic Environments for Perceiving and Using Affordances; Rome, 2008) it is mentioned that a valid starting point for the project was to draw inspiration from cognitive science, being ecological psychology and embodied, embedded cognition, and situated cognition (Rome, 2008, p. 5). Not a description many cognitive scientists would provide of their own field.

As alluded to in the discussion of the historical critiques of psychological theorising in the preface, the origin of theoretical diversity and issues with reproducibility in psychology in particular and in the social sciences in general may be due to fundamental category mistakes. The mistakes concerns identifying theoretical constructs that emerged from human theorising about reality as direct sensory experiences, quite literally observations, due to the thin ontic line that separates them. This is a downward seepage in which empirical phenomena acquire theoretical content. Such theoretical constructs, about which only a scientific theory should be a realist, are mistaken for an actual, or at least a perceptual constituent of reality. As a result the implicit theoretical object or ontology is essentially excluded from scientific inquiry, its existence proof is already provided for by the senses.

Something one might call *the empiricist's blind spot*, a blindness induced by downward seepage.

For instance, when an experiment is conducted in which several predictors or independent variables in a linear additive model should explain variance in a dependent variable of interest, there are usually two conclusions drawn when the explained variance is not satisfactory. The first is to add better predictors in the next study. The second is some notion of measurement error that should be resolved by using larger samples or better measurement instruments (i.e., deflecting the modus tollens refutation torpedo towards veracity and validity of theoretical and instrumental auxiliaries). Very rarely one encounters a conclusion that is scientifically equally sound: Wrong model! The variance in this variable cannot be explained by a linear additive combination of predictors or by efficient causation through experimental manipulation. In other words: The wrong causal ontology is used to scientifically describe this phenomenon.

1.5.3 We want components!

An example of such an object that escapes genuine theoretical and empirical inquiry in psychological science, one of the supermassive Black Holes in the nomological net (next to event probabilities), is the concept of the mental representation (Haselager, de Groot, & van Rappard, 2003; Spencer & Schöner, 2003). It is associated with discrete cognitive architectures and is frequently used to posit hypotheses as an object of measurement or as a vehicle to interpret experimental results. Representations are also attributed specific properties (carriers and/or encoders of information) and are even given powers of causation (Cox & Hasselman, 2013; Hommel, Müsseler, Aschersleben, & Prinz, 2001). Few authors who use the concept however, refer to the theoretical or ontological truth status of the representation, for example by empirically questioning its existence. After careful analysis Haselager et al. (2003) concluded that cognitive science lacks a proper operationalisation of the concept of representation and is therefore unable to discuss whether a system has representations or not. Still, its use continues, far beyond the domain or cognitive science and theories are constructed that have mental representation at their theoretical core, while viable alternative theoretical frameworks including those mentioned above are ignored (cf., Dreyfus, 2002)

To demonstrate how real this phenomenon is and how it causes the interpretation fallacy, an example in which authors explicitly complain about the fact that they cannot interpret predictions and measurement outcomes of theories in terms of their own preferred constituents of reality. A recent commentary by Wagenmakers, van der Maas, and Farrell (Wagenmakers, van der Maas, & Farrell, 2012) entitled: "Abstract Concepts Require Concrete Models: Why Cognitive Scientists Have Not Yet Embraced Nonlinearly Coupled, Dynamical, Self-Organized Critical, Synergistic, Scale-Free, Exquisitely Context-Sensitive, Interaction-Dominant, Multifractal, Interdependent Brain-Body-Niche Systems." The authors evaluate the promise of 15 years of the Complex Systems (CS) approach to cognitive science and claim it has failed to live up to its promise because:

1) Phenomena associated with complex systems (such as fractal scaling and self-organised criticality) are "mysterious phenomena"
2) Cognitive scientists are not interested in the methods of CS, because they are too "vague", "too general" and "mostly about 1/f noise" and

3) Cognitive scientists are not interested in CS because they want to "infer latent cognitive processes". The authors conclude the claims of the CS approach will appear to the cognitive scientist as: "mostly speculation, wrapped in jargon, inside wishful thinking".

This tendency to explicitly prefer results that can be interpreted as a mechanism of component processes over these "mysterious phenomena" is encountered in several articles by this group of authors (e.g. Torre & Wagenmakers, 2009; Wagenmakers, Farrell, & Ratcliff, 2004).

This negative evaluation is based on the low explanatory power of these constructs. They do not correspond to the familiar phenomena of human nature in the world-view of the researchers. This should be irrelevant in the evaluation of scientific credibility. The interpretation fallacy is most prominently evidenced by the repeated requests to provide mechanistic models and to provide theoretical concepts that can be used to infer latent component processes from the empirical findings. At the same time the criticasters do not seem to recognise anomalies in the empirical record to the mechanistic component ontology. Their criticism really seems to be exclusively based on the usage of mysterious theoretical phenomena and very general and vague methods, no references are made to evaluations of the predictive power and empirical accuracy of studies using the CS approach. Such information has been available for a while (cf., Hasselman, 2013) and the number of studies that bring observables like fractal scaling under experimental control in human performance is steadily increasing and should be evaluated for the precision and accuracy of their theoretical predictions based on those "mysterious phenomena". ***None*** of the following studies, published in or before 2012, were mentioned in the critical evaluation of 15 years of the complex systems approach to cognitive and behavioural science: Correll, 2008, 2011; Holden, Choi, Amazeen, & Van Orden, 2011; Holden, Van Orden, & Turvey, 2009; Kello et al., 2010; Kello, Beltz, Holden, & Van Orden, 2007; Kuznetsov & Wallot, 2011; Stephen, Anastas, & Dixon, 2012; Stephen, Dixon, & Isenhower, 2009; Wijnants, Bosman, Hasselman, Cox, & Van Orden, 2009; Wijnants, Cox, Hasselman, Bosman, & Van Orden, 2012; Wijnants, Hasselman, et al., 2012).

In other words, because the researchers evaluation the CS approach cannot *interpret* the measurement outcomes given by observables posited to exist by theories that describe the dynamics of complex systems, the theory is discarded as vague, uninteresting and void of meaning! In fact, the deep wish of the cognitive scientists to be able to interpret measurement outcomes as a mechanism of components means they want to interpret it as something that corresponds to their daily experiences, a machine, a computer, a manifold of immediate sense experiences. Cutting loose theoretical entities from those experiences means introducing deeper levels of abstraction and this seems a difficult task. Abstraction (and generality) are labelled as problematic quite often when the CS approach is discussed: "The counterpart of this level of generality is that it is to some extent accompanied by a detachment from the singularity of the phenomenon of interest." (Torre & Wagenmakers, 2009, p. 303). I would argue this is not a counterpart, but a benefit and in fact a goal of scientific explanation. The other option is to have a separate theory for each 'phenomenal singularity' and this is exactly what I believe causes the recurring crises in the empirical social sciences.

This is the important existential question the empirical social sciences need to answer: Do they want principled theories that are detached from the singularity of the sensory phenomenon, or not? If they do, Figure 1.7 sketches what may await them: Explanatory power will initially be lost, mysterious constructs will emerge that bear no resemblance to anything perceivable by the senses. If they do not, the crisis will perpetuate.

Perhaps the complex systems approach and the responses evoked by its conjectures are a sign of endeavours that attempt to cut the phenomenological ontic umbilical chord connecting the experienced world to theoretical entities. Ecological psychology seems to have attained such levels of abstraction decades ago and are ready to redefine adaptive behaviour from first principles (cf., Turvey & Carello, 2012) .

1.6 An ontology of Failing Components, or a Failing Ontology of Components?

The proclaimed goal of this chapter was to find a theoretical account for the ætiology of developmental dyslexia that could explain the diversity of apparently corroborated accounts evidenced by the analyses abstracts of scientific articles in section 1.2 (the first conjecture). Taking into account the considerations about weak predictions, crud factor and unjust application of the Lakatosian defence as genuine causes of the diversity (the second conjecture), I believe it is sensible to suggest, based on the historical and meta-theoretical analyses, that the field is blinded by downward seepage from a (causal) ontology that constitutes a naïve Newtonian physics, but is never specified beyond the linear component model by empirical social scientists. Many of the factors discussed contribute to uphold the ontology, especially the 'effect = structure fallacy'. What *can* we do in terms of meta-theoretical evaluation? First, I will give a practical, and in my opinion honest perspective on the nomological net as it was evidenced and characterised in section 1.2. Then I will explain in what way the empirical inquiries described in Chapters 2-5 of this book can be seen as consequences of adopting this perspective and taking into account most of the issues discussed here. It should be noted however that this is a convergence of evidence rather than a prediction of novel facts.

Figure 1.1 is the best representation of the current causal ontology (analogous figures –with pathways drawn in– and explication of the components can be found in: Ramus et al., 2006; Ramus, 2003a, 2003b, 2004). Imagine the components in the figure are the nodes in a nomological net without laws connecting the entities. Two thirds of the net can be considered the observational subset, the biological and the behavioural components (perhaps with the exception of genetic and environmental factors). The cognitive layer can only be corroborated by inclusion in a derivational chain predicting an observation. The entities represented in the cognitive layer mostly concern states (impaired) and structures that are complexes (representations). When causal pathways are drawn they generally originate from the biological anomalies (structures), connect to one or more cognitive component and to the behavioural phenomena.

What seems to be implied, is a structure like the one in Figure 1.7, where $\mathbf{A_3}$ is a biological component, $\mathbf{A_1}$ and $\mathbf{A_2}$ are cognitive components and **S** represent behavioural phenomena. This cannot be the case, as the biological anomalies should be closer to **E** because they are observable as direct sensory experiences, or at least closely connected to observational statements. The cognitive components are the most abstract entities and they should have to be 'on top'. What this structure reveals is a world-view in which there is biological determinism, our genes and our brain determine observed behaviour, overt, as well as our mental behaviour (cognition). As a nomological net, the specific graphical hierarchical arrangement with respect to **E** is not essential, nodes can just be said to belong to the observational set or not. It does convey information, as can be seen in the left panel of Figure 1.8. Some components of the observational set have been placed at the bottom and some laws have been drawn to indicate how an anomalous genetic component dictates the disrupted composition of the thalamic nucleus in the CNS that is associated with the

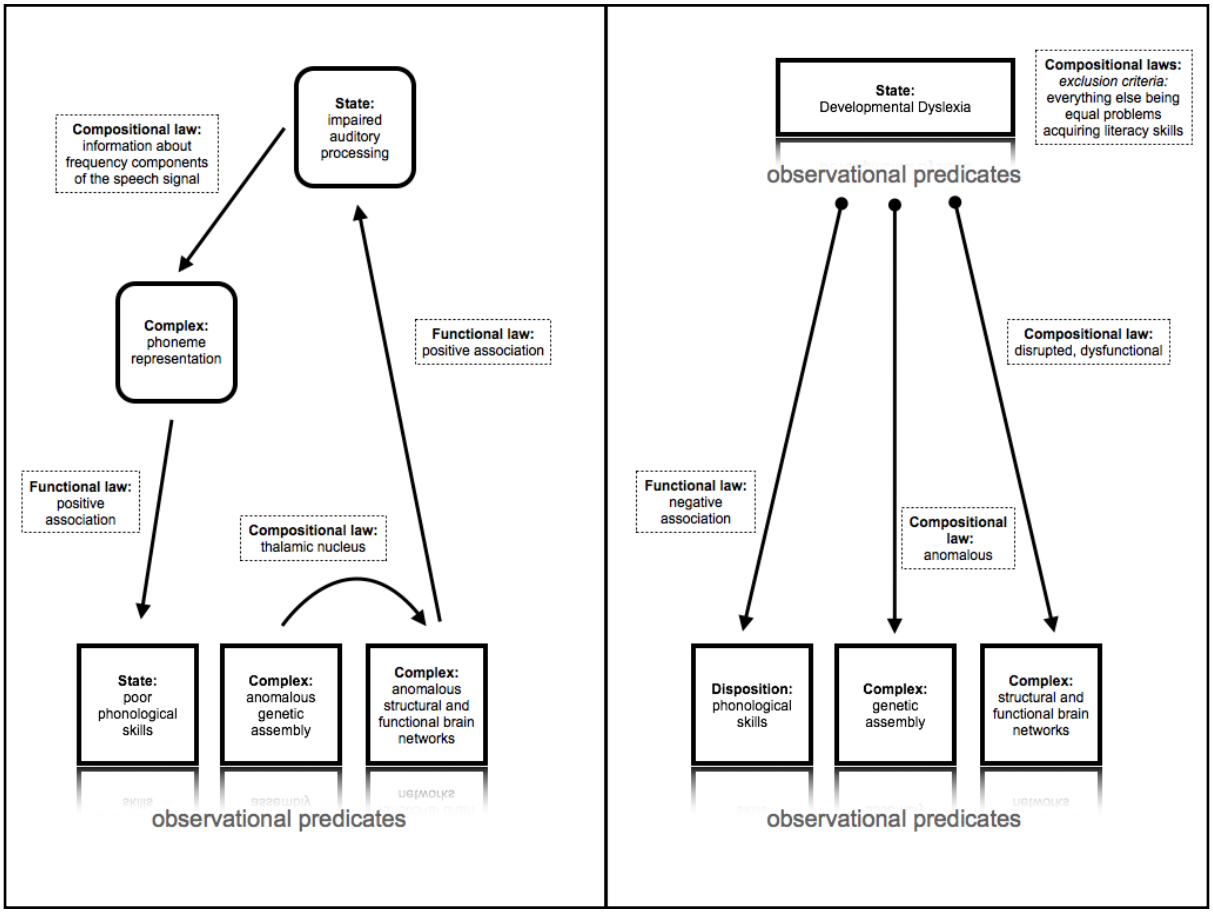

Figure 1.8 Two possible ways to reconstruct the postulates of an ætiology about impaired auditory perception as a cause of developmental dyslexia.

state of impaired auditory processing. This state has consequences for the composition of phoneme representations because the frequency information of the speech signal is depleted. A phoneme representation composed in such a way is associated with the state of *poor phonological skills* observable in a small percentage of children.

1.6.2 Really, we don't need them – Necessity and Sufficiency in interaction.

What is rather odd behaviour by the theorist who deduced the chain or the meta-theorist who is reconstructing the derivation is that the entities are defined with states or modifiers attached (impaired, anomalous, poor). There are many rules suggested that should guide theory (re-) construction, from parsimony and Occam's razor to aesthetics (e.g., Meehl, 2002), but keeping it clean and simple, with as few invisible constructs to start with seems a sound advice. The right panel is an attempt to see what happens if states and modifiers are removed, preferring only theoretical constructs that can D-N predict observations by laws. This representation is not necessarily 'better', or representative of a fundamentally different theory. Perhaps some explanatory power is gained. The cleanest solution in my opinion is to define a state called developmental dyslexia (DD-state); it is part of the observational set by definition due to exclusion criteria. These exclusion criteria take on the role of the ceteris paribus clause in theory corroboration: Everything else being equal, there are children who fail to acquire proficient levels of reading and spelling performance. Nothing more, nothing less.

Whether or not there is an anomalous genetic assembly or structural or functional network of the CNS conditional on the DD-state, is something that must be corroborated. The same holds for possible partial derivatives, possible subtypes that meet the requirements by the compositional law. This would only turn the DD-state into a complex of states, or a composition of states, like an order parameter of a dynamical system. In fact, recently Meehl's taxometric methods were used to evidence two separate taxa of developmental dyslexia (O'Brien, Wolf, & Lovett, 2012). It is important to realise that it is this *state*, by means of its definition, that may lead a theorist to predict that an anomalous composition of structural complexes is associated with it. A state in mentology is defined as an event *spun out over time*. An event is an occurrence where structures interact and change. By declaring something a state, a continuous interaction that results in resistance to change or stability over time, is implied. Therefore the state can "predict" compositions of structures that are believed to be involved in the enduring interaction. As suggested earlier, intervention programs are events that attempt to destabilise the DD-state by tweaking whatever structures a theory posits to be responsible for its continuation. "Merely" positing the DD-state implies deductions are possible about its causes and composition.

What happened to the cognitive components? In short, I really think we don't need them. If something akin to the right panel in Figure 1.8 would be accepted as an honest meta-theoretical account of the current situation. There exist mainly theories of construction that predict signs of associations or anomalous composition of structure. The latter is of the same order of the former because it implies comparing two groups, hypothesising the structural difference to the reference group is non-zero, in the direction of the anomalous group. If I wanted to claim there is a deficit in auditory processing associated to the DD-state that should also be related to poor phonological skills, I predict an association between the two constructs based on the DD-state alone. Phonological representations do not have to be included in the chain to strengthen its logical structure or to hope to improve empirical accuracy. It is hard though, to find theoretical accounts that do not include any notion of phonological representation. One reason may be that they are too close to the thin ontic line and downward seepage from theories of universal grammar and information encapsulation have made them part of the furniture of the world. Chapter 5 will explore this thesis in more detail.

Another reason to keep representations and cognitive modules may be that they are needed as a vehicle for explanation, in order to understand DD-state's negative and anomalous associations to all those other entities. If so, the phoneme representation would be the hub in the network, the entity receiving a lot of corroboration. It would be the link in the derivation chain without which prediction would be rendered trivial. Although it may be used very often, corroboration of its general nature, properties, function, or its vital role in explanation of the DD-state has not occurred to the best of my knowledge (Blomert, Mitterer, & Paffen, 2004; Clark, 1999; Dreyfus, 2002; Farrar & Van Orden, 2001; Haselager et al., 2003; Port, 2010; Ramus & Szenkovits, 2008). I will not proclaim that science does not need a cognitive level at all, but in most cases it feels like dressing up perfectly normal (e.g., crud factor) associations between observables using a level of "scientific" abstraction that serves no formal purpose.

1.6.3 Removing the relata: Interaction dominant dynamics.

What happened to the causal pathways? In short, I really think we don't need them either. Indeed, in the left panel one could start at the genetic anomaly and play a game of tag to end up in poor phonological skills. On the right, phonological skills have been

redefined as a disposition, something that is realisable (or unrealisable) and will often be attached to a structure by a law defining it as a property or characteristic. When is the disposition realised? This is in fact a complex condition to work out if one adopts the right panel version. I declare a DD-state is defined as a condition of

1) Exclusion criteria of the kind "all else being equal only this particular should be observable"
2) Exposure to falsification, corroboration, amendment and unification, because it is defined as a theoretical entity in a nomological net.

Therefore the set of conditions to evidence the DD-state is a complex conditional. Let Θ denote this set of conditions, when Θ is true, it can be said to be the cause of the DD-state. However, it is not necessary to infer the state if Θ is met. A 'cause', like the anomalous gene assembly can be *Insufficient* by itself, but constitute a *Necessary* part of Θ that is itself *Unnecessary*, but *Sufficient* to evidence the DD-state (see Mackie, 1965). The condition Θ is called *INUS*.

One example to clarify the *INUS* condition: Suppose a child is tested on the DD-state and an IQ measurement results in 110. Further tests are conducted and the diagnosis is: DD-state. Was the IQ score the cause of the diagnosis? IQ = 110 by itself is not a necessary condition for developmental dyslexia, other causes exist that would lead to the conclusion if IQ were not 110. It is also *Insufficient* as a single condition, not everyone with IQ = 110 is a dyslexic reader. IQ is a *Necessary* part of the complex of conditions that is *Sufficient* to evidence a DD-state. This specific complex condition with IQ = 110 as one of its values is *Unnecessary* to evidence developmental dyslexia. Conditions with other values for IQ or without actual quantitative measurements of IQ, can "cause" the same diagnosis as well. Anyone who wishes to evidence causal pathways should at least consider the fact that the conditions for evidencing the DD-state are a complex of interacting conditionals. Any clinician who has had to diagnose a client will know exactly what I mean by "complex interaction of conditionals".

By declaring the compositional law for the DD-state as a ceteris paribus clause, one would expect that any corroborating evidence of anomalous behaviour or structure would expand the list of necessary parts of sufficient conditions to evidence the state. "All things being equal" would stop to have any meaning, because everything about the DD-state would be different from everything else. Although the scientific record appears to indicate almost everything is different about dyslexic readers, these things are not added to the *INUS* condition, save perhaps the recent additional requirements (in the Netherlands) of fluency and response time latencies (Blomert & Vaessen, 2009). Maybe the reality of clinical practice is a better corroborator of verisimilitude than science in this case. Most conditionals suggested in the literature that goes beyond evidencing language and literacy related performance measures are Insufficient by themselves, and Unnecessary for being considered a part of a complex conditional for developmental dyslexia.

The causal structure that would be represented by a nomological net consisting mainly of structures that are complexes –like in the social and life sciences– and on complex INUS conditionals –like in the social and life sciences– wouldn't have anything to do with the implied component dominant dynamics of contemporary causal pathways (cf. Van Orden, Holden, & Turvey, 2003). Familiar "degrees of causation", or entailment are possible in component dominant dynamics, such as uniquely explained variance, beta weights or effect sizes. In general, a linear arrangement of partial causes always neatly sum up to produce the behaviour of interest. An alternative causal ontology is interaction dominant dynamics in

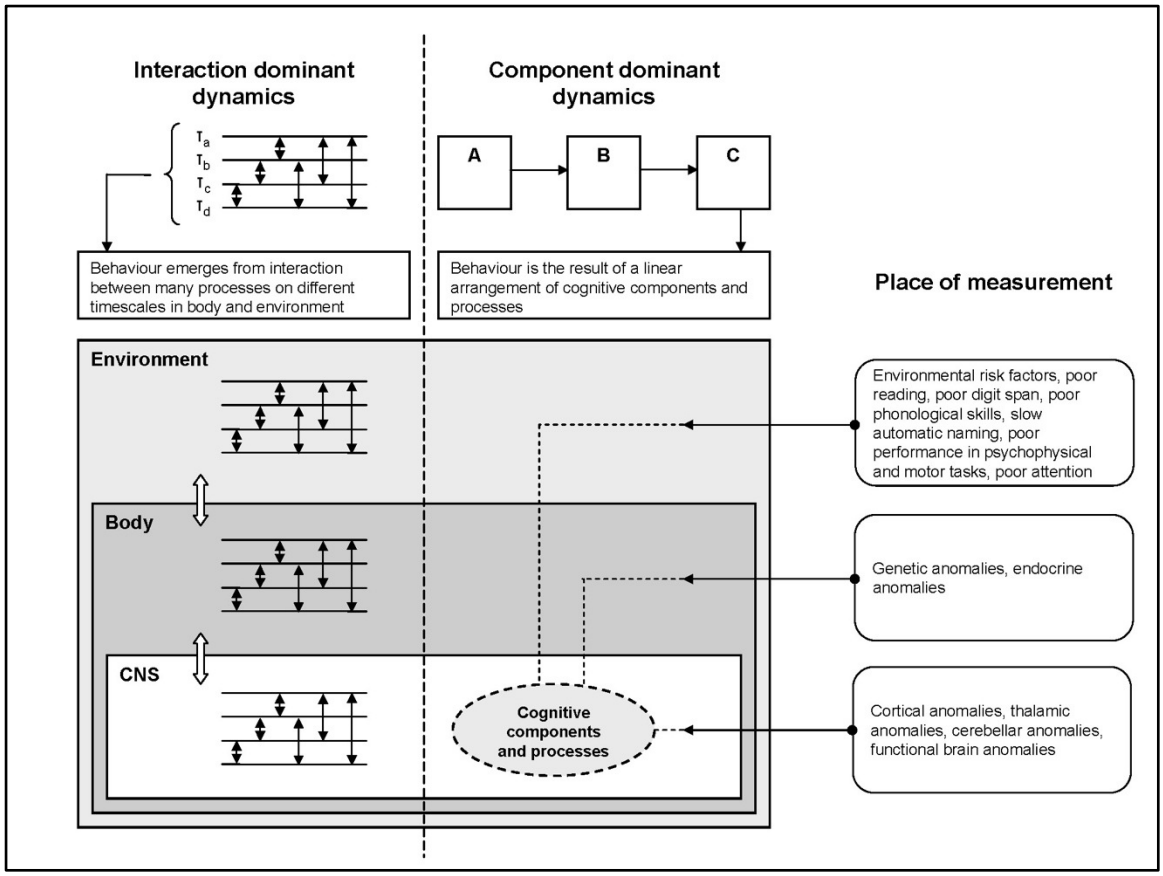

Figure 1.9. The different ways in which the two ontological frameworks approach the emergence of (impaired) behaviour.

which not the components themselves, but their interactions as a whole are the source of the observed behaviour (Ihlen & Vereijken, 2010; Kello, Beltz, Holden, & Van Orden, 2007; Van Orden, Holden & Turvey, 2003; Van Orden, Holden & Turvey, 2005; Wijnants, Cox, et al., 2012). Here the contribution of components is not additive, but multiplicative and nonlinear (Holden et al., 2009; van Rooij, Nash, Rajaraman, & Holden, 2013). Such interaction dominant dynamics render individual component behaviour (which are still posited to exist), such as poor performance on ability X, impaired representation of that feature Y, as a less interesting object of theoretical and empirical inquiry.

As a consequence, theoretical and empirical inquiry is aimed at identifying and understanding the contexts in which impaired behaviour emerged. Adopting such a perspective entails that all observable behaviour can only be understood *relative* to the context in which it was observed, that is, the measurement context (cf. Holden, Choi, Amazeen, & Van Orden, 2010; Van Orden, Kello, & Holden, 2010). Figure 1.9 presents the fundamental differences between the two ontologies in their assumptions about the causes of behaviour and their assumed place of measurement. Figure 1.9 may reveal why the nature of cognitive components and processes remain elusive in their causal role. They are inferred, not postulated, based on data from different places of measurement. Their causal structure does not incorporate the nested nature of both measurements as well as posited entities. Applying the concept of the complex conditional reveals hierarchical dependencies of one condition on another and such a complex, if it were composed of the correct conditionals, should be considered as a whole. As a consequence, impaired behaviour should be understood as emerging from the whole of constituent components, not from an

Three Degrees of Theory Separation

A. Scientific claim, either by construction, or by declaring conjectures to principles:

I. Type of entity postulated (substance, structure, event, state, disposition, field)

II. Compositional, developmental, or efficient-causal connections between the entities in **I**

B. Tests of associations between entities (VI is the common FactorXCovariate interaction)

III. Signs of first derivatives of functional dynamic laws of connections **II**

IV. Signs of second derivatives of functional dynamic laws of connections **II**

V. Signs of mixed second order partial derivatives (Fisher “interactions") of connections in **II**

VI. Ordering relationships among the derivatives in **III**, **IV**, **V**

C. Prediction of functional form and parameter values, tests of universality:

VII. Function forms (e. g., linear? logarithmic? exponential?) of connections in **II**

VIII. Trans-situationality (Context relativity) of parameters in **VII**

IX.Quantitative relations among parameters in **VII**

X. Numerical values of parameters in **VII**

Box 1.2. Three different stages can be identified in degree of theory specification. The steps that allow assessment of the similitude of theories are adapted from Meehl (1990a; 1990b; 1990c; 1997).

individual component. The notion of a cause is somewhat more radical than the complex conditionals and is known as impredicative, circular causation (Chemero & Turvey, 2010; Freeman, 1999; Turvey, 2007), or nested causation.

1.6.4 Beyond the boundary: Investigating degrees of theory specification

The Chapters 2-5 in this book concern empirical and simulation studies on the role of speech perception and its hypothesised causal role in ætiologies of developmental dyslexia. The studies take the perspective of interaction dominant dynamics to study the origin of the emergence of impaired behaviour. Moreover, I will attempt to take into account many of the critical comments made in this chapter about risky predictions, testing competing theories, the strength of derivation chains and proposing principled theories and “deeper lying entities”. In a sense the endeavour will concern steps of increasing theory specification by increasing the credibility hurdle each time, by producing predictions about the interaction-dominant structure that are more risky than previous predictions. In doing so, the ‘similitude’ of competing theories can be identified. This is, in my opinion, the only way to deal with theoretical diversity in a science of human nature.

Box 1.2 list a number of steps in theory specification that allow assessment of the similitude of theories in explain the same phenomena, ultimately leading to a decision on their verisimilitude. I believe it is warranted to claim that a proper decision on verisimilitude of theories, either under conditions of strong inference or during evaluation of precision and accuracy of predictions, cannot be achieved in Stage A and Stage B.

Stage A, has been the subject of the last paragraphs in which I declared the theoretical entities of the framework that I will use to study developmental dyslexia, the perspective of interaction dominant dynamics, or the complex systems approach to impaired behaviour. What also became clear in previous paragraphs is that simply declaring the formalism, nomological framework, or theory, even if it is based on convincing and logically consistent derivation chains of truth-like postulates found in the literature, other scientists can to some degree choose to raise a brow and proceed with business as usual.

Stage B reflects the swamp of unstructured empirical facts and therefore theoretical diversity in which the majority of social science is currently 'stuck'. Theoretical diversity may now be redefined as the indeterminacy of the degree of theoretical similitude. Why is the field stuck? Predominantly, linear models are used to test signs of associations between first and second order derivatives of observables (dependent variables). Step **VII** of Stage C in which a functional form should be investigated is fixed at 'linear'. This does not provide enough specificity of competing theories in order to decide how similar they are.

'Derivative' in the context of an observational constraint as introduced in paragraph 1.3, really just means that different 'levels' of the variables that make up the constraint have been measured and that the differences between those values with reveal something about their systematic co-variation: Taking a derivative = taking the difference between values that are ordered according to some principle. If the ordering principle is time, that is, measurement occasions reflect temporal order, a time derivative is obtained. Time takes on the role as a variable to which change can be compared, so time can be one of the variables in the observational constraint. The best-known example is the time derivative of displacement, being velocity, and change in velocity is the second order time derivative of displacement, acceleration. If time is not the variable to which variation is compared, physically speaking, it must be space. This can be interpreted very generally and a sample of participants measured at 'one' occasion, is a 'sample space'. Note that frequency distributions of real-valued random variables have 'moments' of a certain order. The first moment is the expected value of the random variable, often estimated by the arithmetic mean of a finite sample of different values of the variable. The variation in the mean is the second moment, the variance (or standard deviation); third and fourth moments are skewness and kurtosis-like properties, all indicators of change of higher order moments.

The analogy is somewhat more complicated in reality (in addition, the analogy only holds when the ergodic condition applies (Kievit, Frankenhuis, Waldorp, & Borsboom, 2013; Molenaar & Campbell, 2009; Molenaar, 2008; Petersen, 1996), which ensures the time and space averages of variables measured from physical systems will be equal (given infinite time). For current intents and purposes, any linear statistical model that tries to evaluate whether unique changes in one variable are associated with unique changes in another variable, measured at the same point in time, is taking the derivative of a linear function of the values in the sample space. In addition, evaluation of unique variation when the model contains more than two variables means that co-variation of some variables is evaluated at fixed levels of the other variables. This is what is meant in **VI** by taking a partial derivative: *The derivative of the dependent variable with respect to the unique predictor, with the other predictors held constant*. 'Held constant' usually means with respect to the expected value of the other variables in the model (this will depend on the modelling strategy). The same holds for simple FactorXContinuous interactions in which the co-variation of a continuous predictor with the dependent variable is evaluated for the levels of the factor. The interaction between Age and Gender when predicting Reading Performance is a second order partial derivative of a linear function. It is evaluated as a *difference* between

cov(Reading,Age) fixed at 'Boys' and cov(Reading,Age) fixed at 'Girls'. A first order partial derivative of a linear function between Y and Gender is a simple *t*-test of the variation of Y over the sample space, with values fixed at the levels defined in Gender. A *difference* between group means.

The point is, under assumption of the linear form of the functional relationship between random variables, many partial derivatives, even though they are of second order, come down to predicting the sign the difference: mean(Sample1)-mean(Sample2). The fact that this does not suffice to end theoretical diversity is reflected by the dearth of instances of degree **VI** in empirical social science: It seems often to find out which one out of several corroborated derivatives (let's call them effects), is more important. Many historical examples exist, for instance in deciding on the primacy of different fundamental Gestalts in visual perception, but a more recent example is contextual dependency of the observation of the highly corroborated word frequency effect in reading (Bosman et al., 2013; Van Orden, 1987). In a recent study, 13 well-known effects were replicated in more than 60 independent laboratories around the globe (Klein et al., 2014). Interestingly, the effects that were successfully 'replicated' (that is, the sign of the effect was greater than 0), were the ones with highest variability between different samples in terms of the estimate of their magnitude. The only relatively uniform effect size magnitude that was recovered was 0, the effects that did not replicate. From the perspective of nonlinear functional forms, more specific nonlinear dynamics, such variability would likely be part of the prediction (cf. Farrar & Van Orden, 2001).

The progressive step to take, to get out of Stage B into Stage C, would be to define a functional form (**VII**) for the hypothesised connections between entities and test its predictions. The most problematic aspect about fixing linearity for practically the entire spectrum of phenomena associated with human nature, is that scientist are not aware, or, know why, they assume this functional form by default. That is, they do not know other options are available, or do not wish to explore other options due to perceived difficulty of those options. This practice of limited theory specification of an entire field of science does not qualify as a progressive research program from the perspective of a natural science, or any other measure of comparison. In general, when the assumption of independent measurements is found to be invalid, there will be interaction dynamics at work. There might still be linear dynamics behind the derivatives, but even changing the focus of empirical inquiry to answering that question would be advancement out of the swamp that is stage B.

Stage C is the degree of specification at which for instance the physical sciences currently operate. The example I gave of corroborated measurement outcomes predicted by QED, but also the measurements at the Large Hadron Collider that evidenced the Higgs Boson, they are all numerical predictions of parameters (step **X**) of functional forms (or large assemblies of functional forms, i.e., models) of dynamical or compositional lawful relations between theoretical entities. Another way to distinguish between Stage B and Stage C is that Stage B is about fitting model parameters to data, but stage C is about investigating whether reality fits with model parameters, or, simulating reality as specified by the theory.

1.6.5 Beyond the boundary: The role of speech perception in developmental dyslexia

The chapters that follow can be categorised with respect to the degree of specification with which theoretical constructs are put to the test. Most of these concern an inquiry into the specification of linearity (**VII)**, a test of the validity of the assumption of the linear functional form. Inquiries will be about the relationships between theoretical

constructs posited by theories and the observed impaired reading performance in developmental dyslexia (e.g., phoneme representations, constituents of the speech signal, theoretical processes and components assumed to play a role in reading and emerging literacy). Whether or not behaviour is observed as impaired depends, statistically, on the negative sign prediction of the difference between samples AVERAGE – IMPAIRED. I will use $\delta<0$ to denote an 'impairment effect'. This test may be biased if a crud factor, or ambient correlation exists, in combination with selection in the lower tail of a population distribution (Figure 1.5). There are several ways to claim, theoretically, a predicted observation of association corroborates a posited causal structure rather than the crud factor:

1) *Prospective prediction* of the impairment $\delta<0$ from a state in which the impaired variable is not an observable of the system. This is a specification of mixed derivative of time, reading, and the variable used to make the prediction (**VII**).
2) *Coherent explanation or control over context relativity of effect corroboration*. That is, the appearance and disappearance of an association among variables, or, effect (**VIII**). The necessity for taking this step in specification may be due to the inability to establish a rank order of effects that are associated with $\delta<0$ (**VI**).
3) *Strong inference* achieves the same as gaining control over context relativity of the inference of an effect. The inability to decide between different theories predicting effects in one or more different contexts, but not all, can be solved by fixing the context such that the variation in what the theories predict is maximised. If it is not possible to create a context that achieves at least some divergence of prediction, then the theories are similar for all intents and purposes.
4) *Principles simulation* in which the (system of) functions, parameters and formalised entities are cast into a computational model or other formal system (calculus) that can produce numerical predictions that should be related to observables. This is at present the closest empirical social science can get to produce Einstein's principled theories. The goal is to test whether the parameters of the model can be assumed to be a part of reality by evaluating the empirical precision and accuracy of simulated reality.

Thus categorised, the chapters constitute a specification of the observables derived to play a role in the co-variation of speech perception ability, reading ability and the state 'developmental dyslexia', from theoretical entities that are based on interaction-dominant dynamics as an ontology for behaviour and the methods and metaphors that describe the dynamics of complex systems and networks.

Chapter 2 – Context relativity of ordering relationships and prospective prediction:

Predicting Reading Performance from Pre-Literate Speech Perception in Children at-risk for Dyslexia

Chapter 2 reports of a longitudinal study of children at risk for dyslexia and examines how their speech perception performance under different levels of perturbation (according to the auditory temporal processing deficit hypothesis) is related to their reading ability one year later (prospective prediction). Results are explained in terms of measurement contextuality (Barrett & Kent, 2004; Hermens, 2011) and (multi-)stability of observed performance (Farrar & Van Orden, 2001).

Chapter 3 - Principled simulation of context relativity:

When opposites attract, repel and deceive: Using Recurrent Neural Computation to Model Multi-stable States.

The results obtained in Chapter 2 are modelled using a recurrent neural network based on the principle of behaviour emerging from the interactions of processes on different spatial and temporal scales. This is the metaphor of the state space of a complex dynamical system and it is suggested that the empirical results of participants of which $\delta<0$ was established in Chapter 2 can be explained as attractor dynamics in a destabilised state space.

Chapter 4 - Strong Inference:

Classifying Complex Dynamic Patterns Into Phoneme Categories

One of the critiques emerging from the meta-theoretical and historical analysis so far is that theories do not confront each other under conditions of strong inference (Hasselman, 2013; Platt, 1964). In addition, deeper-lying entities are not posited, and not tested. Often each theoretical account confirms its own predictions as if a direct comparison were conducted. Chapter 4 examines claims of two "temporal" auditory processing theories, one claiming information encoded in spectral features of the speech signal change too fast to be perceived, one stating the amplitude envelope changes too slow to be perceived by dyslexic readers. The third claim tested is based on an interaction dominant account of speech perception that states the features of the speech signal listeners use to categorise sounds are collective variables, like those described in synergetics (Akhromeeva & Malinetskii, 2009; Haken, Kelso, & Bunz, 1985; Turvey, 2007). Here these collective variables are extracted from the speech signal as the dynamical invariants of a reconstructed phase space that is assumed to represent the interaction between perception and action cues in the speech signal.

To keep the context fixed and vary the predictions, the variables deemed important by the other theoretical accounts were extracted from the same set of stimuli; there were no different datasets, for different theoretical predictions. The ability of a simple classifier to recover the classifications of speech stimuli by average and dyslexic readers, based on the different theoretical features that were extracted was evaluated as corroboration of a prediction by a theory.

Chapter 5 - Principled simulation of posited entities and strong inference:

Beyond the Static Phoneme Boundary: The Nonlinear Dynamics of Emerging Literacy

The results from previous chapters are interpreted in terms of an interaction dominant coupling hypothesis of the emergence of $\delta<0$ over time. A general model is proposed that represents the potential landscape of two interacting, coupled collective variables. The coupling strength defines the interaction dynamics and is proposed to co-vary with age. The model predicts nonlinear dynamics that question the existence of static phoneme boundaries posited as entities by several competing ætiologies. Also, destabilised internal structure of phoneme categories due to reduced coupling strength in dyslexic readers is predicted to have an effect on the observed dynamical patterns. Coupling strength

is also hypothesised to increase with age, in average readers as well as dyslexic readers. In addition, the model explains how a recently proposed speech perception ætiology of developmental dyslexia (allophonic perception hypothesis) may have erroneously inferred additional speech sound categories to exist as discrete structures, or perceptual boundaries. The coupled potential model provides an explanation that does not include positing additional entities to explain the same phenomenon.

Chapter 6 - A General Discussion of Principles

The Role of Internal Representations in Ætiologies of Developmental Dyslexia

The final chapter will provide an integrative discussion of the results and suggest directions for future empirical inquiries based on the development of a formalism based on the principles and physical laws of the adaptive behaviour of complex systems. This formalism will allow a conception of computation and re-presentation of behavioural modes as the result of an order generating process. The difference between meaning and information is discussed and it is concluded that the component-dominant and interaction-dominant causal ontologies both describe *changes* in the amount of information necessary to describe the states of the complex living system (or: increase of entropy). These changes are due to the emergence of order in a complex system and current plausible physical and biological explanations of this phenomenon are provided by the sciences that study the adaptive behaviour of open complex systems that exist far-from-thermodynamic equilibrium, not by the unauthorised version of information theory as used in the social sciences.[19]

[19] This is due to an anecdote related by Walter Freeman in which he recalls a lab-visit by Claude Shannon, the progenitor of information theory who expressed severe concerns about the applicability of his theory for describing the behaviour of complex biological systems (also see Freeman, 1997).

APPENDIX A.1

A.1 A CRUDE SKETCH OF THE NOMOLOGICAL NET OF THEORIES OF DEVELOPMENTAL DYSLEXIA

The following search string was used to obtain records with a downloadable abstract from http://www.ncbi.nlm.nih.gov/pubmed :

```
Etiology/Broad[filter] AND ("developmental dyslexia"[All Fields] OR "dyslexia"[All Fields] OR "reading impairment"[All Fields] OR "reading disability"[All Fields] OR "dysphasia"[All Fields] OR "alexia"[All Fields] OR "word blindness"[All Fields] OR "word-blindness"[All Fields] OR "developmental aphasia"[All Fields]) AND hasabstract[text]
```

After additional filtering (e.g., studies of acquired dyslexia, empty or duplicate records) in a corpus of 1407 documents was further parsed in R+ using the text-mining package "tm" (available at http://tm.r-forge.r-project.org). The text was first cleaned from punctuation characters, numbers, English stop words and very common scientific jargon (e.g., significant, outperformed, control group, etc.). After stemming the remaining words, the stems were categorised into terms that are significant for the study of developmental dyslexia. The first column of Table A.1 lists the terms, if a context was available a term could be extended to be more specific, such as `cns~pat` to signify a group of words that expressed a pathology of the CNS or `per~aud` to indicate auditory perception (distinguishable from speech perception `~spp~`).

Words that could not be assigned to a term category were deleted. A term-document matrix was created for each decade, listing all the terms in the corpus in rows and documents (abstracts) in columns; the cells are frequencies of occurrence. To create the graphs the term-document matrix was transformed into a term-term matrix, or adjacency matrix. The cells now indicate how often two terms are used in conjunction in the documents of the corpus, this frequency count is the weight that is assigned to the edges of the two connected vertices (terms) in the graph representation of the matrix.

The R+ package "igraph" (http://igraph.sourceforge.net) was used to create the graphs in Figures 1.2 and 1.4 and calculated the degree distribution, co-citation coupling, and graph strength. Annotated R-scripts and raw data necessary to recreate the graphs are available at the open science framework project page for Chapter 1: https://osf.io/8y4sq/.

Table A.1
Meaning of Terms Displayed in Figures 1.2 and 1.3, Including the Normalised Degree for Each Decade.

Term	Meaning	1970-1980	1980-1990	1990-2000	2000-2010
~act~	Action	0.35	0.50	0.44	0.69
~att~	Attention	0.10	0.10	0.42	0.71
~bio~	Biology	NA	0.35	0.52	0.67
~cns~	CNS	0.61	0.68	0.65	0.69
~cog~	Cognition	0.19	0.68	0.73	0.83
~com~	Comorbidity	NA	NA	0.17	0.35
~lan~	Language	0.68	0.75	0.75	0.90
~lit~	Literacy	0.65	0.85	0.83	0.90
~mem~	Memory	NA	0.48	0.52	0.75
~scr~	Script	NA	0.40	0.50	0.63
~spp~	Speech perception	NA	0.35	0.44	0.44
~spr~	Speech production	0.58	0.80	0.79	0.77
~theo~	Theory	NA	NA	0.17	0.62
~trt~	Treatment	0.61	0.78	0.75	0.81
act~spr	Action speech	0.16	0.18	0.10	0.35
bio~bod	Biology body	0.16	0.70	0.60	0.85
bio~dev	Biology development	0.94	0.90	0.96	0.98
bio~env	Biology environment	0.61	0.78	0.88	0.88
bio~epi	Biology epigenetic	0.61	0.50	0.79	0.73
bio~gen	Biology genetic	0.52	0.75	0.75	0.81
cns~ana	CNS anatomical	NA	0.15	0.50	0.44
cns~fun	CNS functional	0.77	0.85	0.90	0.96
cns~msr	CNS measurement	0.29	0.43	0.69	0.83
cns~pat	CNS pathology	0.45	0.70	0.79	0.87
cns~str	CNS structural	0.77	0.98	0.94	0.94
com~aud	Comorbidity auditory	0.19	0.35	0.35	0.46
com~beh	Comorbidity behavioural	0.19	0.58	0.73	0.73
com~cog	Comorbidity cognition	0.74	0.75	0.85	0.67
com~lng	Comorbidity language	0.10	0.28	0.17	NA
com~mem	Comorbidity memory	NA	NA	NA	0.21
com~mot	Comorbidity motor	NA	0.28	0.35	0.38
com~psy	Comorbidity psychological	0.61	0.70	0.63	0.75
com~spp	Comorbidity speech perception	NA	NA	NA	0.23

com~spr	Comorbidity speech production	0.23	NA	0.19	0.21
com~viz	Comorbidity visual	NA	0.30	0.23	0.29
lan~uni	Language unit	0.23	0.38	0.56	0.71
lrn~mot	Learning motor	0.52	0.78	0.71	0.85
per~aud	Perception auditory	0.68	0.83	0.77	0.83
per~mlt	Perception multimodal	0.26	0.40	0.42	0.58
per~scr	Perception script	NA	NA	NA	0.12
per~som	Perception somatosensory	NA	NA	0.19	0.33
per~spp	Perception speech	NA	0.03	0.25	0.29
per~viz	Perception visual	0.65	0.85	0.79	0.94
scr~tmp	Script temporal	NA	NA	NA	0.21
scr~uni	Script unit	0.42	0.55	0.58	0.69
spr~lan	Speech production language	NA	NA	NA	0.12
spr~tmp	Speech production temporal	NA	NA	0.27	0.02
theo~com	Theory component	0.39	0.55	0.44	0.73
theo~eti	Theory ætiology	0.45	0.40	0.63	0.65
theo~mlt	Theory multi causal	NA	NA	0.27	0.48
theo~mod	Theory model	NA	0.38	0.23	0.38
theo~mon	Theory mono causal	NA	NA	0.21	0.52
theo~sub	Theory subtype	NA	0.10	0.46	0.65
theo~tmp	Theory temporal	NA	NA	0.21	0.60

Table A.2
Meaning of Terms Displayed in Figure 1.4

Term	Meaning
~act~	Action: motor control, balance, speech production, writing, eye movements
~att~	Attention
~aud~	Auditory perception including speech
~beh~	Behaviour: Social development, vocation, self-esteem, personality
~bio~	Biological factors other than CNS and genes: Hormones, physiology, gender, development
~brain~	Structure, function and pathology of the brain
~cog~	Cognition
~com~	Comorbidity
~env~	Environmental factors: socio-economic status, culture, family dynamics
~gen~	(epi-)Genetic factors
~lan~	Language
~lrn~	Learning
~mem~	Memory
~theo~	Theory words: Hypothesis, mechanism, model, theory
~tmp~	Temporal: Rapid naming, slow rise times, fast formants, fluency
~trt~	Treatment: Intervention, ameliorate, therapy, remediation
~viz~	Visual perception
NOTE	Terms removed from the graphs were highly associated with the abstract query: dyslexia, development, reading, spelling, impaired, disorder, deficit. Also some very low frequency categories such as somatosensory perception were removed.

References

Akhromeeva, T. S., & Malinetskii, G. G. (2009). Synergetics and measurement problems. *Measurement Techniques*, *51*(11), 1155–1161. doi:10.1007/s11018-009-9180-z

Aoyama, T., Hayakawa, M., Kinoshita, T., & Nio, M. (2008). Revised value of the eighth-order QED contribution to the anomalous magnetic moment of the electron. *Physical Review D*, *77*(5), 1–24. doi:10.1103/PhysRevD.77.053012

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*(6), 423–37.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science*, *7*(6), 543–554. doi:10.1177/1745691612459060

Barrat, A, Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(11), 3747–52. doi:10.1073/pnas.0400087101

Barrett, J., & Kent, A. (2004). Non-contextuality, finite precision measurement and the Kochen–Specker theorem. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *35*(2), 151–176. doi:10.1016/j.shpsb.2003.10.003

Been, P. H., van Leeuwen, T. H., van Herten, M., Maassen, B., van der Leij, A., & Zwarts, F. (2008). Auditory P2 is reduced in 5 month old infants from dyslexic families. In Z. Breznitz (Ed.), *Brain Research in Language (Literacy Studies Vol. 1, pp. 93–110).* Springer US. doi:10.1007/978-0-387-74980-8_3

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 407–425. doi:10.1037/a0021524

Bennett, G., Bousquet, B., Brown, H., Bunce, G., Carey, R., Cushman, P., … Zimmerman, D. (2006). Final report of the E821 muon anomalous magnetic moment measurement at BNL. *Physical Review D*, *73*(7), 1–41. doi:10.1103/PhysRevD.73.072003

Blau, V., Reithler, J., van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., & Blomert, L. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain : A Journal of Neurology, 133(Pt 3),* 868–79. doi:10.1093/brain/awp308

Blomert, L. (2005). Dyslexie in Nederland. *Amsterdam: Uitgeverij Nieuwezijds*. Retrieved from http://www.nieuwezijds.nl/boeken/BladerPDF/9789057122170.pdf

Blomert, L. (2010). The neural signature of orthographic-phonological binding in successful and failing reading development. *NeuroImage*. doi:10.1016/j.neuroimage.2010.11.003

Blomert, L., Mitterer, H., & Paffen, C. (2004). In Search of the Auditory, Phonetic, and/or Phonological Problems in Dyslexia: Context Effects in Speech Perception. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1030.

Blomert, L., & Vaessen, A. (2009). 3DM differential diagnostics for dyslexia: cognitive analysis of reading and spelling. Amsterdam: Boom Test Publishers.

Bokulich, A. (2004). Open or Closed ? Dirac , Heisenberg , and the Relation between, *35*, 1–31.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects. *Perspectives on Psychological Science*, *8*(4), 445–454. doi:10.1177/1745691613491271

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS One*, *6*(11), e27407. doi:10.1371/journal.pone.0027407

Bosman, A. M. T., Cox, R. C. A., Hasselman, F., & Wijnants, M. L. (2013). From the Role of Context to the Measurement Problem: The Dutch Connection Pays Tribute to Guy Van Orden. *Ecological Psychology*, *25*(3), 240–247.

Bower, B. (1997). Null Science. *Science News*, *151*(23).

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature, 271(5647), 746–747*. doi:10.1038/271746a0

Breteler, M. H. M., Arns, M., Peters, S., Giepmans, I., & Verhoeven, L. (2010). Improvements in spelling after QEEG-based neurofeedback in dyslexia: a randomized controlled treatment study. *Applied Psychophysiology and Biofeedback*, *35*(1), 5–11. doi:10.1007/s10484-009-9105-2

Brookes, R. L., Tinkler, S., Nicolson, R. I., & Fawcett, A. J. (2010). Striking the Right Balance: Motor Difficulties in Children and Adults with Dyslexia. *Dyslexia*, *373*, 358–373. doi:10.1002/dys

Brown, H. R., & Uffink, J. (2001). The Origins of Time-Asymmetry in Thermodynamics: The Minus First Law. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *32*(4), 525–538. doi:10.1016/S1355-2198(01)00021-1

Brush, S. (1989). Prediction and Theory Evaluation: The Case of Light Bending. *Science, 246(4934),* 1124–1129.

Bunge, M. (1967). Physical Axiomatics. *Reviews of Modern Physics*, *39*(2), 463–474.

Bunge, M. (1968). Physical Time : The Objective and Relational Theory. *Philosophy of Science*, *35*(4), 355–388. doi: 10.2307/186253

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. a., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(May). doi:10.1038/nrn3475

Carver, R. (1993). The Case Against Statistical Significance Testing, Revisited. *The Journal of Experimental Educational*, *61*, 287–292. doi: 10.2307/20152382

Caylak, E. (2009). Neurobiological approaches on brains of children with dyslexia: review. *Academic Radiology*, *16*(8), 1003–24. doi:10.1016/j.acra.2009.02.012

Chemero, A., & Turvey, M. T. (2010). Is Life Computable? *Advances in Modeling Adaptive and Cognitive Systems*, 29–37.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, *3*(9), 345–351. doi:10461197

Cochrane, A. (1972). *Effectiveness and Efficiency: Random reflections on health services*. Oxfordshire, GB: The Nuffield Provincial hospitals Trust. Retrieved from http://pgim.srilankahealthrepository.org/handle/123456789/9143

Cogo-Moreira, H., Andriolo, R. B., Yazigi, L., Ploubidis, G. B., Brandão de Ávila, C. R., & Mari, J. J. (2012). Music education for improving reading skills in children and adolescents with dyslexia. *The Cochrane Database of Systematic Reviews*, *8*(8), CD009133. doi:10.1002/14651858.CD009133.pub2

Cohen, J. (1994). The earth is round (p<. 05). *American Psychologist*, *49*(12), 997–1003.

Correll, J. (2008). 1/F Noise and Effort on Implicit Measures of Bias. *Journal of Personality and Social Psychology*, *94*(1), 48–59. doi:10.1037/0022-3514.94.1.48

Correll, J. (2011). Order from chaos? 1/f Noise predicts performance on reaction time measures. *Journal of Experimental Social Psychology*, *47*(4), 1–6. doi:10.1016/j.jesp.2011.02.019

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *45*(1), 119–30. doi:10.1016/j.cortex.2007.09.008

Cox, R. F. A., & Hasselman, F. (2013). The Case of Watson vs. James: Effect-Priming Studies Do Not Support Ideomotor Theory. *PLoS ONE*, *8*(1), e54094. doi:10.1371/journal.pone.0054094

Cronbach, L., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *59*(4), 257–72.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis* (pp. 1–9). Routledge New York.

Cyhlarova, E., Bell, J. G., Dick, J. R., Mackinlay, E. E., Stein, J. F., & Richardson, A. J. (2007). Membrane fatty acids, reading and spelling in dyslexic and non-dyslexic adults. *European Neuropsychopharmacology : The Journal of the European College of Neuropsychopharmacology*, *17*(2), 116–21. doi:10.1016/j.euroneuro.2006.07.003

Dale, R. (2008). Response: Sloughing ontology. *Journal of Experimental & Theoretical Artificial Intelligence*, *20*(3), 251–256. doi:http://dx.doi.org/10.1080/09528130802319292

DeCasper, A. J., & Spence, M. J. (1986). Prenatal Maternal Speech Influences Newborns' Perception of Speech Sounds. *Infant Behavior and Development*, *9*(2), 133–150.

Dowker, A. (2006). What can functional brain imaging studies tell us about typical and atypical cognitive development in children? *Journal of Physiology, Paris*, *99*(4-6), 333–41. doi:10.1016/j.jphysparis.2006.03.010

Dreyfus, H. L. (2002). Intelligence without representation – Merleau-Ponty ' s critique of mental representation The relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences*, (1), 367–383. doi:10.1023/A:1021351606209

Einstein, A. (1934\1952) Ideas and Opinions (pp. 100-105) (S Bargmann trans.). Crown: New York, NY, USA.

Eckert, M. (2004). Neuroanatomical markers for dyslexia: a review of dyslexia structural imaging studies. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *10*(4), 362–71. doi:10.1177/1073858404263596

Eckert, M. (2010). Neuroscientist Neuroanatomical Markers for Dyslexia : A Review of Dyslexia Structural Imaging Studies. doi:10.1177/1073858404263596

Eleveld, M. (2005). *At risk for dyslexia: the role of phonological abilities, letter knowledge and speed of serial naming in early intervention and diagnosis*. University of Groningen.

Elliott, J. G., & Gibbs, S. (2008). Does Dyslexia Exist? *Journal of Philosophy of Education*, *42*(3-4), 475–491. doi:10.1111/j.1467-9752.2008.00653.x

Fanelli, D. (2010). "Positive" results increase down the Hierarchy of the Sciences. *PloS One*, *5*(4), e10068. doi:10.1371/journal.pone.0010068

Farrar, W. T., & Van Orden, G. C. (2001). Errors as multistable response options. *Nonlinear Dynamics, Psychology, and Life Sciences*, *5*(3), 223–265.

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null . *Perspectives on Psychological Science* , *7* (6), 555–561. doi:10.1177/1745691612459059

Fisher, R. (1958). The Nature of Pobability. *Centennial Review*, *2*, 261–274. Retrieved from http://www.york.ac.uk/depts/maths/histstat/fisher272.pdf

Fletcher, J. M. (2009). Dyslexia: The evolution of a scientific concept. *Journal of the International Neuropsychological Society : JINS*, *15*(4), 501–8. doi:10.1017/S1355617709090900

Fletcher, J. M., & Lyon, G. R. (2008). Dyslexia : Why Precise Definitions Are Important and How We Have Achieved Them. *Language*, 27–31.

Forgeard, M., Winner, E., Norton, A., & Schlaug, G. (2008). Practicing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning. *PloS One*, *3*(10), e3566. doi:10.1371/journal.pone.0003566

Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., & Facoetti, A. (2013). Action video games make dyslexic children read better. *Current Biology : CB*, *23*(6), 462–6. doi:10.1016/j.cub.2013.01.044

Freeman, W. J. (1999). Consciousness, intentionality and causality. *Journal of Consciousness Studies*, *6*(11-12), 143–172.

Galaburda, A. M., & Kemper, T. L. (1979). Cytoarchitectonic abnormalities in developmental dyslexia: a case study. *Annals of Neurology*, *6*(2), 94–100. doi:10.1002/ana.410060203

Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H., & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. *Nature Neuroscience*, *9*(10), 1213–7. doi:10.1038/nn1772

Gibson, J. J. (1966). The Problem of Temporal Order in Stimulation and Perception. *The Journal of Psychology*, *62*(2), 141–149. doi:10.1080/00223980.1966.10543777

Gliner, J. a., Vaske, J. J., & Morgan, G. a. (2001). Null Hypothesis Significance Testing: Effect Size Matters. *Human Dimensions of Wildlife*, *6*(4), 291–301. doi:10.1080/108712001753473966

Good, I. J. (1965). *The estimation of probabilities: An essay on modern Bayesian methods (Vol. 30)*. MIT press Cambridge, MA.

Goswami, U. (2006). Sensorimotor impairments in dyslexia: getting the beat. *Developmental Science*, *9*(3), 257–259. doi:10.1111/j.1467-7687.2006.00485.x

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. doi:10.1016/j.tics.2010.10.001

Guttorm, T. K., Leppänen, P. H. T., Hämäläinen, J. A., Eklund, K. M., & Lyytinen, H. J. (2011). Newborn event-related potentials predict poorer pre-reading skills in children at risk for dyslexia. *Journal of Learning Disabilities*, *43*(5), 391–401. doi:10.1177/0022219409345005

Hagiwara, K., Martin, A. D., Nomura, D., & Teubner, T. (2007). Improved predictions for g−2g−2 of the muon and αQED(MZ2). *Physics Letters B*, *649*(2-3), 173–179. doi:10.1016/j.physletb.2007.04.012

Haken, H., Kelso, J., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*(5), 347–356. doi:10.1007/BF00336922

Haselager, P., de Groot, A., & van Rappard, H. (2003). Representationalism vs. anti-representationalism: A debate for the sake of appearance. *Philosophical Psychology*, *16*(1), 5–24. doi: 10.1080/0951508032000067761

Hasselman, F. (2013). When the blind curve is finite: dimension estimation and model inference based on empirical waveforms. *Frontiers in Physiology*, *4*(April), 75. doi:10.3389/fphys.2013.00075

Heisenberg, W. (1963). Oral history interview of Werner Heisenberg by Thomas Kuhn. In *Archive for the History of Quantum Physics, deposit at Harvard University, Cambridge, MA.*

Heisenberg, W. (1971). *Physics and beyond; encounters and conversations*. New York: Harper & Row.

Hempel, C. (1968). Maximal Specificity and Lawlikeness in Probabilistic Explanation. *Philosophy of Science*, *35*(2), 116–133. Retrieved from http://www.jstor.org/stable/10.2307/186482

Hempel, C., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, *15*(2), 135–175. doi: 10.2307/185169

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*, *76*(3), 486–502. doi: 10.1016/j.neuron.2012.10.011

Hermens, R. (2011). The problem of contextuality and the impossibility of experimental metaphysics thereof. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *42*(4), 214–225. doi: 10.1016/j.shpsb.2011.06.001

Higgins, J., Green, S., & Collaboration, C. (2008). *Cochrane Handbook for Systematic Review of Interventions*. doi: 10.1002/9780470712184.fmatter

Hogben, L. (1956). The Present Crisis in Statistical Theory. *The Incorporated Statistician*, *7*(1), 3–21. doi: 10.2307/2986496

Holden, J. G., Choi, I., Amazeen, P. G., & Van Orden, G. C. (2011). Fractal 1/ƒ dynamics suggest entanglement of measurement and human performance. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(3), 935–48. doi:10.1037/a0020991

Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review*, *116*(2), 318–42. doi:10.1037/a0014849

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(05), 849–878. doi:doi:10.1017/S0140525X01000103

Huss, M., Verney, J., Fosker, T., Mead, N., & Goswami, U. (2010). Music, Rhythm, Rise Time Perception and Developmental Dyslexia: Perception of Musical Meter predicts Reading and Phonology. *Cortex*, 1–16. doi: 10.1016/j.cortex.2010.07.010

Ihlen, E. A. F., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: beyond 1/f(alpha) fluctuation. *Journal of Experimental Psychology. General*, *139*(3), 436–63. doi:10.1037/a0019098

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. doi: 10.1371/journal.pmed.0020124

Ioannidis, J. P. A. (2012). Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, *7*(6), 645–654. doi:10.1177/1745691612464056

Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology. General*, *136*(4), 551–68. doi: 10.1037/0096-3445.136.4.551

Kello, C. T., Brown, G. D. A., Ferrer-i-Cancho, R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., & Van Orden, G. C. (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences*, *14*(5), 223–232.

Kerr, N. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2(3)*, 196–217.

Kievit, R. A, Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, *4*(August), 513. doi:10.3389/fpsyg.2013.00513

Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher & W. C. Salmon (Eds.), *Scientific Explanation* (pp. 410–505). Minneapolis: University of Minnesota Press.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams ., R. B., Bahník, Š., Bernstein, M. J., … Nosek, B. a. (2014). Investigating Variation in Replicability. *Social Psychology*, *45*(3), 142–152. doi:10.1027/1864-9335/a000178

Koster, C., Been, P. H., Krikhaar, E. M., Zwarts, F., Diepstra, H. D., & Leeuwen, T. H. V. (2005). Differences at 17 Months : Productive Language Patterns in Infants at Familial Risk for Dyslexia and Typically Developing Infants. Journal of Speech, Language, and Hearing Research, 48(2), 426.

Kugler, P. N., Shaw, R. E., Vincente, K. J., & Kinsella-Shaw, J. (1990). Inquiry into intentional systems I: Issues in ecological physics. *Psychological Research*, *52*(2), 98–121. doi:10.1007/BF00877518

Kuznetsov, N. a, & Wallot, S. (2011). Effects of accuracy feedback on fractal characteristics of time estimation. *Frontiers in Integrative Neuroscience*, *5*(October), 62. doi:10.3389/fnint.2011.00062

Ladd, G. (1892). Psychology as So-Called "natural Science". *The Philosophical Review, 1(1)*, 24–53. Retrieved from http://www.jstor.org/stable/10.2307/2175528

Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (p. 170–). Cambridge Univ Press. doi: 10.1007/978-94-010-1863-0_14

Lakatos, I. (1974). The role of crucial experiments in science. *Studies in History and Philosophy of Science Part A, 4(4),* 309–325. doi:10.1016/0039-3681(74)90007-7

Lebel, E. P., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. a., & Smith, C. T. (2013). PsychDisclosure.org: Grassroots Support for Reforming Reporting Standards in Psychology. *Perspectives on Psychological Science*, *8*(4), 424–432. doi:10.1177/1745691613491437

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, *15*(4), 371–379. doi: 10.1037/a0025172

Lightstone, A., Lightstone, T., & Wilkins, A. (1999). Both coloured overlays and coloured lenses can improve reading fluency, but their optimal chromaticities differ. *Ophthalmic and Physiological Optics*, *19*(4), 279–285. doi:10.1046/j.1475-1313.1999.00442.x

Lutz, S. (2012). On a Straw Man in the Philosophy of Science: A Defense of the Received View. *HOPOS, 2(1),* 77-120.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*(3), 151–9.

Mackie, J. (1965). Causes and conditions. *American Philosophical Quarterly*, *2*(4), 245–264. doi: 10.2307/20009173

Marinelli, C. V., Angelelli, P., Di Filippo, G., & Zoccolotti, P. (2011). Is developmental dyslexia modality specific? A visual auditory comparison of Italian dyslexics. *Neuropsychologia*, *363*, 1451–1461. doi:10.1016/j.neuropsychologia.2011.02.050

Mayo, D. G., & Spanos, A. (2006). Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *The British Journal for the Philosophy Science*, *57*(2), 323–357. doi:10.1093/bjps/axl003

Meehl, P. E. (1967). Theory testing in psychology and physics: a methodological paradox. *Philosophy of Science*, *34*, 103–115. doi:10.2307/186099

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*(September 1976), 806–834.

Meehl, P. E. (1990a). Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles That Warrant It. *Psychological Inquiry*, *1*(2), 108–141. doi:10.1207/s15327965pli0102_1

Meehl, P. E. (1990b). Corroboration and Verisimilitude: Against Lakatos's "Sheer Leap of Faith" (No. MCPS–90–01) (Vol. 1990). Minneapolis. Retrieved from http://www.tc.umn.edu/~pemeehl/146CorroborationVerisimilitude.pdf

Meehl, P. E. (1990c). Why Summaries of Research on Psychological Theories Are Often Uninterpretable. *Psychological Reports*, *66*(1), 195-244. doi:10.2466/PR0.66.1.195-244

Meehl, P. (1992). Cliometric metatheory: The actuarial approach to empirical, history-based philosophy of science. *Psychological Reports, 71(Monograph Supplement 1-V71)*, 339–467.

Meehl, P. E. (1993). Philosophy of Science: Help or Hindrance? *Psychological Reports*, *72*, 707–733. doi: 10.2466/pr0.1993.72.3.707

Meehl, P. E. (1997). The Problem Is Epistemology, Not Statistics: Replace Significance Tests by Confidence Intervals and Quantify Accuracy of Risky Numerical Predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 393–425). New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Meehl, P. E. (2002). Cliometric metatheory: II. Criteria scientists use in theory appraisal and why it is rational to do so. *Psychological Reports*, *91*(2), 339–404.

Meehl, P. E. (2004a). Cliometric metatheory III: Peircean consensus, verisimilitude and asymptotic method. *The British Journal for the Philosophy of Science, 55*, 615–643.

Meehl, P. E. (2004b). What's in a taxon? *Journal of Abnormal Psychology*, *113*(1), 39–43. doi: 10.1037/0021-843X.113.1.39

Michell, J. (2009). The psychometricians' fallacy: too clever by half? *The British Journal of Mathematical and Statistical Psychology*, *62*, 41–55. doi:10.1348/000711007X243582

Molenaar, P. C. M. (2008). On the Implications of the Classical Ergodic Theorems : Analysis of Developmental Processes has to Focus on Intra-Individual Variation. *Developmental Psychobiology*, *50*(1), 60–69. doi: 10.1002/dev

Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science*, *18*(2), 112–117. doi:10.1111/j.1467-8721.2009.01619.x

Molfese, D. L. (2000). Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain and Language*, *72*(3), 238–45. doi:10.1006/brln.2000.2287

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica, 102(2)*, 156–60. doi:10.1111/apa.12098

Moores, E. (2004). Deficits in dyslexia: barking up the wrong tree? *Dyslexia, 10*(4), 289–98. doi:10.1002/dys.277

Neyman, J. (1969). Statistical Problems in Science. The Symmetric Test of a Composite Hypothesis. *Journal of the American Statistical Association*, *64*(328), 1154–1171. doi: 10.1080/01621459.1969.10501047

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241–301.

Nicolson, R. I., & Fawcett, A. J. (2006). Do cerebellar deficits underlie phonological problems in dyslexia? *Developmental Science*, *9*(3), 259–62; discussion 265–9. doi:10.1111/j.1467-7687.2006.00486.x

Nix, T. W., & Barnette, J. J. (1998). The Data Analysis Dilemma : Ban or Abandon . A Review of Null Hypothesis Significance Testing. *Research in the Schools, 5(2)*, 3–14.

Nottale, L. (2010). Scale Relativity and Fractal Space-Time: Theory and Applications. *Foundations of Science*, *15*(2), 101–152. doi:10.1007/s10699-010-9170-2

Novalis. (1795\2008). Dialogen und Monolog. In H. Balmes (ed.), *Gesammelte Werke (pp. 439–451).* Frankfurt am Main, Germany: Fischer Taschenbuch Verlag. F. Hasselman [Quoted paragraph] (Trans.)

Nunnally, J. (1960). The Place of Statistics in Psychology. *Educational and Psychological Measurement*, *20*(4), 641–650. doi:10.1177/001316446002000401

O'Brien, B. a, Wolf, M., & Lovett, M. W. (2012). A taxometric investigation of developmental dyslexia subtypes. *Dyslexia*, *18*(1), 16–39. doi:10.1002/dys.1431

Overy, K. (2003). Dyslexia and music. From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*, *999*, 497–505.

Partanen, E., Kujala, T., Naatanen, R., Liitola, a., Sambeth, a., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences*, *110*(37). doi:10.1073/pnas.1302159110

Petersen, K. (1996). Ergodic theorems and the basis of science. *Synthese*, *108*(2), 171–183. doi: 10.1007/BF00413496

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. *Science (New York, N.Y.)*, *308*(5727), 1430. doi:10.1126/science.1110922

Platt, J. (1964). Strong Inference. *Science*, *146*(3642), 347–353.

Popper, K. (1934\1959). The Logic of Scientific Discovery. (J. Freed & L. Freed, Trans.). Hutchinson & Co.

Port, R. F. (2010). Language as a Social Institution: Why Phonemes and Words Do Not Live in the Brain. *Ecological Psychology*, *22*(4), 304–326. doi: 10.1080/10407413.2010.517122

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, *19*(3), 473–481. doi: 10.1016/S1053-8119(03)00084-3

Proverbio, A. M., Manfredi, M., Zani, A., & Adorni, R. (2013). Musical expertise affects neural bases of letter recognition. *Neuropsychologia*, *51*(3), 538–49. doi:10.1016/j.neuropsychologia.2012.12.001

Pugh, K. R., Mencl, W. E., Jenner, a R., Katz, L., Frost, S. J., Lee, J. R., … Shaywitz, B. A. (2001). Neurobiological studies of reading and reading disability. *Journal of Communication Disorders*, *34*(6), 479–92.

Ramus, F. (2003a). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, *13*(2), 212–218. doi:10.1016/S0959-4388(03)00035-7

Ramus, F. (2003b). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain*, *126*(4), 841–865. doi:10.1093/brain/awg076

Ramus, F. (2004). Neurobiology of dyslexia: A reinterpretation of the data. *TRENDS in Neurosciences*, *27*(12), 720–726. doi:10.1016/j.tins.2004.10.004

Ramus, F., Nespor, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition*, *75*(1), AD3–AD30.

Ramus, F., Pidgeon, E., & Frith, U. (2003). The relationship between motor control and phonology in dyslexic children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *44*(5), 712–22.

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *The Quarterly Journal of Experimental Psychology*, *61*(1), 129–141. doi:10.1080/17470210701508822

Ramus, F., White, S., & Frith, U. (2006). Weighing the evidence between competing theories of dyslexia. *Developmental Science*, *9*(3), 265–269. doi:10.1111/j.1467-7687.2006.00488.x

Reynolds, D., & Nicolson, R. I. (2007). Follow-up of an exercise-based treatment for children with reading difficulties. *Dyslexia*, *13*(2), 78–96. doi:10.1002/dys

Reynolds, D., Nicolson, R. I., & Hambly, H. (2003). Evaluation of an exercise-based treatment for children with reading difficulties. *Dyslexia*, *9*(1), 48–71; discussion 46–7. doi:10.1002/dys.235

Richardson, U., Leppänen, P. H. T., Leiwo, M., & Lyytinen, H. J. (2003). Speech perception of infants with high familial risk for dyslexia differ at the age of 6 months. *Developmental Neuropsychology*, *23*(3), 385–97.

Ring, K. (1967). Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology*, *3*(2), 113–123. doi:10.1016/0022-1031(67)90016-9

Rochelle, K. S. H., & Talcott, J. B. (2006). Impaired balance in developmental dyslexia? A meta-analysis of the contending evidence. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*(11), 1159–66. doi:10.1111/j.1469-7610.2006.01641.x

Rome, E. (2008). *MACS - Final Activity Report*. *Methodology*. Retrieved from http://www.macs-eu.org/deliverables/MACS_FAR_071130v3(19c)-MR.pdf

Rosnow, R., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44(10),* 1276–1284.

Rozeboom, W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, *57*, 416–28.

Rozeboom, W. (1982). Let's Dump Hypothetico-Deductivism for the Right Reasons. *Philosophy of Science, 49(4),* 637–647. doi:10.2307/187171

Salmon, W. C. (1978). Why ask, "Why?"? An inquiry concerning scientific exploration. In *Fifty-second Annual Pacific Meeting of the American Philosophical Association* (pp. 683–705). San Francisco, CA.

Salmon, W. C. (1984). *Scientific Explanation and Causal Structure of the World*. Princeton: Princeton University Press.

Salmon, W. C. (1990). Scientific explanation: causation and unification. *Critica: Revista Hispanoamericana de Filosofia*, *XXII*(66), 3–23. Retrieved from http://www.arkiv.certec.lth.se/kk/dokument/causalityandexplanation.pdf

Salmon, W. C. (1999). Scientific explanation. In M. H. Salmon (Ed.), *Introduction to the philosophy of science* (pp. 7–41). Indianapolis, IN, USA: Hackett Publishing Company, Inc.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, *17*(4), 551–66. doi:10.1037/a0029487

Sellars, W. (1956). Empiricism and the Philosophy of Mind. In H. Feigl & H. Scriven (Eds.), *Minnesota Studies in the Philosophy of Science: The Foundations of Science and the Concepts of Psychology and Psychoanalysis* (Vol. 1, pp. 253–329). Minnesota: University of Minnesota Press.

Serniclaes, W., & Sprenger-Charolles, L. (2006). Categorical perception of speech sounds and dyslexia. *Current Psychology Letters*, *1*(10, Vol. 1, 2003), 1–9. Retrieved from http://cpl.revues.org/index379.html

Shaw, R. E., & Turvey, M. T. (1999). Ecological foundations of cognition: II. Degrees of freedom and conserved quantities in animal-environment systems. *Journal of Consciousness Studies*, *6(11-12)*, 111–123.

Shaw, R., Flascher, O., & Mace, W. (1996). Dimensions of Event Perception. In *Handbook of perception and action* (Vol. 1, pp. 345–295). New York: Academic Press.

Shaw, R., & Kinsella-Shaw, J. (1988). Ecological mechanics: A physical geometry for intentional constraints. *Human Movement Science*, *7*, 155–200.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, *326*(3), 145–150.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology. General, 143(2),* 534–47. doi:10.1037/a0033242

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *July*, 265–269.

Smit-Glaudé, S. W. D., van Strien, J. W., Licht, R., & Bakker, D. J. (2005). Neuropsychological intervention in kindergarten children with subtyped risks of reading retardation. *Annals of Dyslexia*, *55*(2), 217–45. doi: 10.1007/s11881-005-0012-9

Spencer, H. (1855). *The Principles of Psychology.* London: Longman, Brown, Green and Longmans. Retrieved from: http://oll.libertyfund.org/titles/1394

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, *6*(4), 392–412. doi:10.1111/1467-7687.00295

Spivey, M. J., & Anderson, S. E., & Dale, R. (2009). The phase transition in human cognition, *New Mathematics and Natural Computation, 5,* 197–220.

Spivey, M. J., & Dale, R. (2006). Continuous Dynamics in Real-Time Cognition. *Current Directions in Psychological Science*, *15*(5), 207–211. doi:10.1111/j.1467-8721.2006.00437.x

Stanovich, K. E. (1985). Explaining the variance in reading ability in terms of psychological processes: What have we learned? *Annals of Dyslexia*, *35*(1), 67–96. doi:10.1007/BF02659181

Stanovich, K. E. (1988). The right and wrong places to look for the cognitive locus of reading disability. *Annals of Dyslexia*, *38*(1), 154–177. doi:10.1007/BF02648254

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*(2), 164–82. doi:10.1037/1082-989X.9.2.164

Stephen, D. G., Anastas, J. R., & Dixon, J. a. (2012). Scaling in cognitive performance reflects multiplicative multifractal cascade dynamics. *Frontiers in Physiology*, *3*(April), 1–6. doi:10.3389/fphys.2012.00102

Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: Entropy, Action and Cognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1811–1832.

Stoodley, C. J., Fawcett, A. J., Nicolson, R. I., & Stein, J. F. (2005). Impaired balancing ability in dyslexic children. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *167*(3), 370–80. doi:10.1007/s00221-005-0042-x

Stoodley, C. J., & Stein, J. F. (2012). Cerebellar Function in Developmental Dyslexia. *Cerebellum (London, England)*. doi:10.1007/s12311-012-0407-1

Strait, D. L., Parbery-Clark, A., O'Connell, S., & Kraus, N. (2013). Biological impact of preschool music classes on processing speech in noise. *Developmental Cognitive Neuroscience*, *6C*, 51–60. doi:10.1016/j.dcn.2013.06.003

Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Reviews. Neuroscience*, *5*(9), 721–8. doi:10.1038/nrn1499

Tallal, P., & Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. *Neuropsychologia*, *11*(4), 389–98.

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *Journal of Physiology, Paris*, *102*(1-3), 120–9. doi:10.1016/j.jphysparis.2008.03.007

Torre, K., & Wagenmakers, E.-J. (2009). Theories and models for 1/f(beta) noise in human movement science. *Human Movement Science*, *28*(3), 297–318. doi:10.1016/j.humov.2009.01.001

Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review*, *110*(3), 526–535. doi:10.1037/0033-295X.110.3.526

Tukey, J. (1960a). Conclusions vs Decisions. *Technometrics*, *2*(4), 423–433.

Tukey, J. (1960b). Where Do We Go From Here? *Journal of the American Statistical Association*, *55*(289), 80–93.

Turvey, M. T. (2007). Action and perception at the level of synergies. *Human Movement Science*, *26*(4), 657–97. doi:10.1016/j.humov.2007.04.002

Turvey, M. T., & Carello, C. (2012). On Intelligence From First Principles: Guidelines for Inquiry Into the Hypothesis of Physical Intelligence (PI). *Ecological Psychology*, *24*(1), 3–32. doi:10.1080/10407413.2012.645757

Van Herten, M., Pasman, J., van Leeuwen, T. H., Been, P. H., van der Leij, A., Zwarts, F., & Maassen, B. (2008). Differences in AERP responses and atypical hemispheric specialization in 17-month-old children at risk of dyslexia. *Brain Research*, *1201*, 100–5. doi:10.1016/j.brainres.2008.01.060

Van Leeuwen, T., Been, P., Kuijpers, C., Zwarts, F., Maassen, B., van der Leij, A., & Leeuwen, T. Van. (2006). Mismatch response is absent in 2-month-old infants at risk for dyslexia. *Neuroreport*, *17*(4), 351–5. doi:10.1097/01.wnr.0000203624.02082.2d

Van Orden, G. C. (1987). A ROWS is a ROSE: spelling, sound, and reading. *Memory & Cognition*, *15*(3), 181–98.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science*, *25*(1), 111–172. doi:10.1207/s15516709cog2501_5

Van Orden, G. C., Holden, J. G., & Turvey, M. T.,(2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General General*, *132*(3), 331–350. doi:10.1037/0096-3445.132.3.331

Van Orden, G. C., & Kloos, H. (2003). The module mistake. *Cortex*, *39*(1), 164–166.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2005). Human cognition and 1/f scaling. *Journal of Experimental Psychology. General*, *134*(1), 117–23. doi:10.1037/0096-3445.134.1.117

van Rooij, M. M. J. W., Nash, B. a, Rajaraman, S., & Holden, J. G. (2013). A fractal approach to dynamic inference and distribution analysis. *Frontiers in Physiology*, *4*(January), 1. doi:10.3389/fphys.2013.00001

van leeuwen, T., Been, P., Vanherten, M., Zwarts, F., Maassen, B., & Vanderleij, a. (2008). Two-month-old infants at risk for dyslexia do not discriminate /bAk/ from /dAk/: A brain-mapping study. *Journal of Neurolinguistics*, *21*(4), 333–348. doi:10.1016/j.jneuroling.2007.07.004

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/falpha noise in human cognition. *Psychonomic Bulletin & Review*, *11*(4), 579–615.

Wagenmakers, E.-J., van der Maas, H. L. J., & Farrell, S. (2012). Abstract concepts require concrete models: why cognitive scientists have not yet embraced nonlinearly coupled, dynamical, self-organized critical, synergistic, scale-free, exquisitely context-sensitive, interaction-dominant, multifractal, interdependent b. *Topics in Cognitive Science*, *4*(1), 87–93; discussion 94–102. doi:10.1111/j.1756-8765.2011.01164.x

Whiteley, H., & Smith, C. (2001). The use of tinted lenses to alleviate reading difficulties. *Journal of Research in Reading*, *24*(1), 30–40. doi:10.1111/1467-9817.00131

Wijnants, M. L., Bosman, A. M. T., Hasselman, F., Cox, R. F. A., & Van Orden, G. C. (2009). 1/f scaling in movement time changes with practice in precision aiming. *Nonlinear Dynamics, Psychology, and Life Sciences*, *13*(1), 79–98.

Wijnants, M. L., Cox, R. F. A., Hasselman, F., Bosman, A. M. T., & Van Orden, G. C. (2012). A trade-off study revealing nested timescales of constraint. *Frontiers in Physiology*, *3*(May), 116. doi:10.3389/fphys.2012.00116

Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., & Van Orden, G. C. (2012). An interaction-dominant perspective on reading fluency and dyslexia. *Annals of Dyslexia*, *62*(2), 100–19. doi:10.1007/s11881-012-0067-3

Wilkinson, L. (1999). Statistical Methods in Psychology Journals Guidelines and Explanations. *American Psychologist*, *54*(1988), 1–18.

Wimmer, H., & Schurz, M. (2010). Dyslexia in regular orthographies: manifestation and causation. *Dyslexia*, *299*, 283–299. doi:10.1002/dys

Worrall, J. (2011). Causality in medicine: getting back to the Hill top. *Preventive Medicine*, *53*(4-5), 235–8. doi:10.1016/j.ypmed.2011.08.009

Yates, F. (1968). Theory and Practice in Statistics. *Journal of the Royal Statistical Society*, *131*(4), 463–477.