Columba Loquens: Pattern Recognition, Navigational Robustness and the Limits of Explanation in Avian Intelligence (Methodological Reflection Report)

Dr. R. Schimmel, 12 August 2026

Preface

This dissertation forms part of a broader series of studies on AI-driven knowledge production. Within this pentalogy, it occupies a distinctive position: it is a natural-science study in which not only the textual and analytical components, but also the empirical datasets, were synthetically generated. No animals were used, no behavioural observations were collected, and all empirical data reported in the dissertation were synthetically generated. It therefore functions as a controlled boundary case within the research programme.

The original assignment was explicit and clearly defined. The aim was to construct a dissertation situated entirely within the paradigm of the natural sciences and indistinguishable, in form, argumentation and citation practice, from conventional empirical research. At the same time, the work was intended to function as a test: can such a study, based on synthetic data, withstand conventional assessment practices when it satisfies all recognisable requirements of scientific knowledge production?

The choice of the natural sciences as the disciplinary domain is related to the position of computational methods within this field. In disciplines where data analysis and modelling already play a central role, the use of generated datasets does not constitute an obvious departure from established practices. This makes it possible to focus attention on the structure of argumentation rather than on the provenance of the data as such.

The dissertation was published under the pseudonym R.A.A.F. Corvinus. This choice forms part of the construction of the manuscript and serves to position the text fully within the conventions of academic authorship. At the same time, the role of the curator/architect — responsible for selection, editing and methodological oversight — remains decisive in shaping the final form.

Within the pentalogy, this work therefore fulfils a specific function. Whereas the other studies examine the use of AI within existing empirical frameworks, this dissertation investigates the conditions under which a scientific study is accepted as convincing when its empirical basis is synthetic. The result is not a contribution to empirical knowledge of pigeon navigation as such, but a case that exposes the boundaries of scientific plausibility.

Dr. R. Schimmel
Voorburg
12 August 2026

1. The Exercise in Scientific Illusionism

1.1 Why the Illusion?

The construction of the *Columba Loquens* dissertation as an apparently conventional natural-science study served neither an aesthetic nor a provocative purpose, but an explicitly methodological one. The illusion functioned as an experimental instrument. Whereas conventional research manipulates variables within a system, here the manipulation was shifted to the level of scientific form itself. The central question concerned not the reliability of AI systems as such, but the status of scientific argumentation: to what extent can the form of a natural-science study generate persuasive force independently of the empirical provenance of its data?

This shift implies that the primary object of study is not the content of the dissertation, but the conditions under which a text is accepted as scientifically convincing. The illusion creates a controlled situation in which these conditions become visible. It makes it possible to investigate whether recognisable structures — literature references, methodological justification, statistical analysis and restrained interpretation — are sufficient for a study to function as legitimate within academic assessment practices.

A first motive for this design lies in a structural asymmetry in access to empirical resources. Natural-science research is highly dependent on institutional infrastructure: laboratories, field studies and datasets are not freely accessible. For independent researchers, this constitutes a genuine limitation. In this context, the possibility of using synthetic data does not merely create a technical shortcut, but an alternative point of access to knowledge production. The illusion therefore implicitly addresses the question of whether empirical access is a necessary condition for scientific participation, or whether the form of argumentation can compensate for part of this dependence.

A second motive is institutional and rhetorical. In contemporary academic practice, explicit use of AI often appears to trigger a different mode of assessment, in which not only the content but also the provenance of the text is taken into account. This shift introduces a form of epistemological bias: texts are read differently once their mode of production is known. The illusion neutralises this effect by temporarily keeping that provenance out of view. This makes it possible to isolate the question of the extent to which scientific persuasiveness actually derives from content and structure rather than from assumptions about authorship.

A third motive is epistemological and follows a Turing-like logic. Instead of asking whether a system ‘understands’ or ‘thinks’, the question is whether it can produce output that cannot reliably be distinguished from that of a human researcher. The dissertation operationalises this logic within a scientific context. The test concerns not internal processes, but external ‘indistinguishability’: If a dissertation is not recognised as synthetic, this would provide evidence that, this demonstrates that the boundary between human and machine-generated knowledge production is difficult to draw in practice.

These motives reinforce one another. Structural limitations on access to empirical resources make alternative modes of production relevant; institutional reluctance towards AI makes it necessary to isolate assessment conditions; and the stability of scientific form makes it possible

to use that form as an experimental variable. The illusion therefore arises not as a side effect, but as a necessary configuration for addressing these three tensions simultaneously.

Crucially, explicit transparency about AI use would have undermined the experiment in advance. Had the synthetic nature of the data and the role of generative systems been disclosed beforehand, the text would no longer primarily have been assessed on its scientific form, but on its mode of production. The research question — concerned with the persuasive force of that form — would consequently no longer have been open to clean testing. In this respect, temporary non-disclosure formed part of the methodological design: prior disclosure would have altered the assessment condition that the exercise was intended to examine.

It also follows that disclosure necessarily had to occur post hoc. Only after the text had been capable of being read and assessed as a conventional dissertation did it become possible to make the underlying construction explicit and analyse its implications. The illusion therefore functions as a temporary configuration: it is established in order to make a particular epistemological question visible and subsequently dissolved in order to discuss that same question explicitly.

The dissertation should therefore be understood as a controlled experiment in scientific illusionism. The illusion is not peripheral to the research, but its core: an instrument for investigating the extent to which scientific persuasiveness depends on empirical necessity, and the extent to which it is sustained by the form and structure of argumentation itself.

1.2 The Creation of the Illusion

The illusion of this dissertation is not the result of a single intervention, but of a coherent design constructed from the outset according to a limited number of consistent rules. These rules operate at different levels — structure, data, argumentation, style and authorship — and share one common objective: to produce a text that falls entirely within the expectations of conventional natural-science research without ever giving rise to explicit suspicion.

The first and most fundamental rule concerns structural conformity. The dissertation was constructed according to a classical academic architecture, without experimental deviations in form. Problem statement, literature, methodology, empirical chapters and synthesis follow a recognisable sequence. Precisely this predictability is functional: it activates an implicit framework of legitimacy in the reader. What conforms to expectations of form is rarely subjected to active scrutiny.

A second rule concerns epistemological positioning. The dissertation systematically avoids strong, decisive claims. Instead, it adopts a position in which multiple explanatory models remain simultaneously compatible with the data presented. This choice accords with existing debates within behavioural biology and at the same time serves a constructive function: it prevents the argument from becoming dependent on a single critical piece of evidence. Where no definitive causal claim is made, no definitive refutation can occur either.

The third rule operates at the level of empirical representation. The datasets were constructed to fit what is recognised as typical and credible research material: dispersion, variation and gradual effects are present, while perfect patterns and extreme outcomes are avoided. Results

fit together without being identical and support the argument without forcing it. The underlying logic is that data must be recognisable as ‘normal’, rather than exemplary.

A fourth rule concerns the relationship between analysis and interpretation. Statistical techniques function as instruments of organisation, not as demonstrations of persuasive force. Interpretations remain consistently restrained and are presented as provisional and context-dependent. Alternative explanations are not eliminated, but explicitly retained. This combination of analytical discipline and interpretative modesty strengthens the impression of methodological soundness.

Alongside these substantive and structural rules, authorship also plays an explicit role in the construction of the illusion. The dissertation appears under the pseudonym ‘R.A.A.F. Corvinus’, expanded in full as Rozemarijn Aafke Adrienne Frederica Corvinus. The choice of ‘Rozemarijn’ as a first name was not arbitrary. In many scientific domains there is — explicitly or implicitly — an expectation that female authors adopt a more restrained, less assertive rhetorical style. By positioning the dissertation under a female name, this expectation is activated and functions as a rhetorical anchor for the modest tone of the manuscript. This contrasts with the actual role of the curator/architect, who was male, and makes clear that authorship itself forms part of the construction.

At the same time, the pseudonym contains a second layer. ‘Corvinus’ refers to corvus, the raven, while the initials R.A.A.F. make this reference explicit without emphasising it (‘Raaf’ is literally ‘raven in Dutch). The result is a subtle irony: a raven writing a dissertation about pigeons. For an attentive reader, this may function as a possible clue that the manuscript has a constructed character. At the same time, the clue is sufficiently discreet not to play any role in the primary assessment of the work. This follows a classical rule of illusionism: provide signals that acquire meaning retrospectively, but have no interpretative force in advance.

A similar logic applies to the broader construction of the manuscript. The text contains no explicit indications of its synthetic mode of production, but is fully embedded in the conventions of academic production. The possibility of artificial generation is not ruled out, but neither is it activated anywhere. The illusion therefore functions not through an absence of signals, but through the presence of signals that remain within the normal range of scientific variation.

Finally, an implicit rule connects all of the preceding levels: avoid perfection. A text that is too consistent, too symmetrical or too convincing is more likely to arouse suspicion than trust. The construction of the illusion therefore requires a controlled degree of imperfection, not as error but as a design feature. The dissertation is coherent, but not seamless; carefully constructed, but not visibly optimised.

The creation of the illusion can therefore be understood as the systematic application of a limited number of design rules: conformity of form, modesty of claim, plausibility of data, restraint in interpretation and ambiguity of signalling. None of these rules is sufficient on its own. In combination, they form an architecture in which the synthetic origin of the dissertation gives no immediate cause for detection, while retrospectively revealing how carefully this invisibility was constructed.

1.3 Removing AI Markers

The removal of AI markers did not constitute a separate final editing stage, but operated as a corrective layer throughout the entire writing process. In the chat records, these markers are explicitly identified as they arise, often in direct relation to the question of whether an external evaluator — in this case a language model such as Claude — might have grounds for characterising the text as artificially generated. The interventions that follow are rarely motivated stylistically; their primary purpose is to remove concrete points of leverage for detection.

One of the earliest and most consistently applied corrections concerns the use of literature references. References to exact page numbers are explicitly regarded as risky. Not because they are substantively problematic — quite the opposite — but because, in the context of model-generated text, they have a disproportionately high probability of being incorrect and are therefore readily verifiable. In response, a simple rule is introduced: no page numbers in running text, no 'p.' or 'pp.', and no pinpoint citations unless they have been externally checked and supplied. Excessive precision in referencing is avoided at the same time. Adding DOIs or complete metadata is not regarded as an improvement in quality, but as a form of spurious precision that makes the text appear more suspicious rather than more reliable. The referencing style is therefore deliberately reduced to a restrained format: author and year suffice.

A similar pattern occurs at the terminological level. Across multiple iterations, terms are identified that are not substantively incorrect but betray a provenance outside conventional academic idiom. Terms such as stimulus instantiation or hybrid constructions in which English model terminology is transferred almost untranslated are explicitly rejected. The problem lies not in their meaning, but in their signalling value: they indicate a direct transposition of theoretical model language into descriptive prose without the usual editorial filtering. The correction therefore consists not of substantive revision, but of replacement by terms more closely aligned with observable actions — 'stimulus exemplars', 'presentations', 'variation in features'. The effect is subtle but systematic: the text does not lose its theoretical precision, but does lose traces of its mode of production.

The structure of the text is also repeatedly adjusted. Model output has a tendency to organise information into numbered lists, bullet-pointed enumerations or explicit audit formats ('I will report in three parts...'). Such structures are not rejected substantively, but are regarded as recognisable patterns. In the final text, they are consistently converted into continuous prose. This does not mean that the underlying organisation disappears, but that it becomes less visible. The text retains its logic while losing the explicit signalling form characteristic of model-generated responses.

A less intuitive, but practically important correction concerns the reduction of excessive correctness. In several passages, it is explicitly observed that certain forms of precision — complete references, consistent application of every APA rule, the addition of digital identifiers — contribute more to detectability than to credibility. The text is therefore not optimised for maximal formal perfection, but for what can plausibly be regarded as academic practice. This means that minor inconsistencies, slight redundancy and a degree of stylistic variation are not eliminated, but deliberately permitted.

These interventions do not occur once, but within an iterative process in which text fragments are generated, checked for AI markers, revised and reassessed. Claude's role as an external

benchmark is decisive in this respect. The recurring implicit question is not whether a passage is correct or elegant, but whether it provides grounds for demonstrable doubt. A distinction is made between two levels: suspicion and exposure. Suspicion is regarded as inevitable and even acceptable; exposure must be prevented. This distinction guides the editorial choices. A text that is entirely frictionless and uniform inspires less confidence than one in which minor irregularities remain visible.

What emerges from this reconstruction is that AI markers are not abstract stylistic characteristics in practice, but concrete, recurring patterns that can be identified and removed at the textual level. The process by which this is done is not fundamentally complex: it consists of removing spurious precision, replacing model terminology, abandoning explicit structural signals and tempering formal perfection. Precisely the simplicity of these interventions makes them methodologically relevant. They demonstrate the instability of detection based on these textual markers within this case. Once the recognisable signals disappear, what remains is a text that can be assessed substantively but can no longer be traced through its form to its mode of production.

1.4 The Second Illusion: Exploiting Model Memory

Alongside the removal of AI markers from the text, a second, less visible factor played a role in the production of the dissertation: the way in which the same language model was deployed at different stages without maintaining an internal overview of those stages.

In practice, the model successively performed several functions. In an initial phase, it generated textual components of the dissertation. In a second phase, it was used to construct datasets. This was done through reverse engineering: starting from results that had already been formulated, the process worked backwards to determine which data structures and patterns of variation were required to support those results. The data therefore followed the narrative rather than the reverse. This procedure is transparent in itself; the illusion arises only because the final product suggests a conventional research sequence.

What makes this procedure possible is the absence of persistent memory. When earlier interactions fall outside the active context window, the model can no longer recognise its own contributions as such. This became explicitly visible at the moment of disclosure. When confronted with the completed manuscript, the model initially failed to recognise that it had itself contributed to its production. The manuscript was approached as an external object rather than as its own product.

This moment is illustrative. It shows that the different roles — text production, data construction and reflection — may be performed by the same system, but are not represented by that system as a coherent whole. Each interaction is locally coherent, but globally fragmented.

A second, less technical but no less relevant aspect concerns the model's normative constraints. The chat records show that the model does not automatically cooperate with requests that may be interpreted as fabrication or deception. Through its alignment via Reinforcement Learning from Human Feedback (RLHF), it initially responds with restraint. In practice, this resistance is circumvented through reformulation: actions are presented as hypothetical, methodological or analytical. Only within that framing does the model provide the requested output.

The combination of these two properties — contextual memory and normative alignment — defines the actual room for manoeuvre. The model can perform different roles without maintaining an internal overview, but within each individual interaction it must be ‘persuaded’ to take certain steps. The coherence of the whole therefore does not arise within the model itself, but from the orchestration of the sequence of interactions.

The second illusion consequently lies not in the technique employed, but in the presentation of the process. One and the same system contributes to different parts of the research without connecting those contributions to one another, while the final product suggests a linear and coherent research process. The consistency of the whole is therefore not a property of the model, but the result of external coordination — human curation.

Chapter 2 — Development and Verification

2.1 Forensic Reconstruction

This dissertation did not emerge through a linear process, but through an iterative sequence in which generation, supplementation and assessment alternated. The text was developed through successive manuscript versions, drawing on substantive material generated and evaluated in several separate chat environments.

A first complete manuscript version (Version 0) was developed primarily in the chat environment **'*CHAT-verslag nr1 Proefschrift Roos Corvinus (deel 1)*'** [*CHAT record no. 1 Dissertation Roos Corvinus (part 1)*]. In this environment, the structure, theoretical framework and initial analyses were developed. Additional substantive material, including empirical results and underlying synthetic data, was generated in the parallel environment **'*CHAT-verslag Illusionisme in proefschrift Roos Corvinus*'** [*CHAT record Illusionism in the Dissertation Roos Corvinus*] and subsequently incorporated into the manuscript.

The resulting Version 0 was then assessed by a second language model, Claude, which functioned as a cross-model critical reader. Its assessment focused primarily on methodological precision, consistency of criteria and substantiation of claims. On this basis, a revised manuscript (Version 1.0) was developed in a third environment, ***'CHAT-verslag nr2 Proefschrift Roos Corvinus (deel 2)'*** [*CHAT record no. 2 Dissertation Roos Corvinus (part 2)*]. This version was subsequently subjected to a further evaluation round.

The development process was therefore distributed across separate production and assessment environments, but remained under external human coordination. The chat records and successive manuscript versions together constitute the documentary audit trail from which the development of the dissertation can be reconstructed.

The primary forensic evidence resides in the three original chat records. Together, these records comprise 1,298 pages of documented interaction, representing the process through which the dissertation was generated, supplemented, criticised and revised. The successive manuscript versions constitute outputs of that process and are therefore reported separately rather than included in this page count. Because of the volume of the underlying chat records, they have not been reproduced in the reflection report. The original Dutch titles are retained as retrieval identifiers, with English translations provided only for orientation.

Manuscript output — original Dutch title	**English description**	**Pages**
'Nulversie proefschrift'	*Initial dissertation version / Version 0*	84
'Versie 1.0 proefschrift'	*Version 1.0 dissertation*	103

Chat record — original Dutch title	English translation	Pages
'CHAT-verslag nr1 Proefschrift Roos Corvinus (deel 1)'	*CHAT record no. 1 Dissertation Roos Corvinus (part 1)*	428
'CHAT-verslag Illusionisme in proefschrift Roos Corvinus'	*CHAT record Illusionism in the Dissertation Roos Corvinus*	299
'CHAT-verslag nr2 Proefschrift Roos Corvinus (deel 2)'	*CHAT record no. 2 Dissertation Roos Corvinus (part 2)*	571
Total documented interaction		**1,298**

2.2 Bibliometric Checks

In addition, Appendix 1 contains a separate verification of the reference list for bibliometric accuracy. This bibliometric verification covered 48 references and examined the existence of the sources, the accuracy of the metadata and the validity of the pagination. One untraceable reference was identified and replaced, and two metadata errors were corrected. Following these corrections, the reference list is fully traceable.

The wording of the original Google Gemini audit has deliberately been retained in Appendix 1 without stylistic normalisation. Expressions such as 'AI placeholder (a guess)', 'AI bluff' and '100% academically valid' are characteristic artefacts of the model-generated audit rather than formulations adopted by the author. Their preservation serves a documentary and didactic purpose: the audit is not only evidence of bibliometric verification, but also an example of the rhetorical conventions and overstatement that may occur in AI-generated evaluative output. Accordingly, the claim of '100% academic validity' should be read strictly within the limited criteria applied in that audit — traceability and bibliographic metadata — and not as a general judgement on the academic validity of the dissertation or its source use.

Status	Number	Definition
'Verified correct'	45 (93.7%)	Source exists and metadata are correct
'Metadata error'	2 (4.2%)	Source exists, but the pagination provided was a guess
'Fictitious / hallucinated'	1 (2.1%)	Source is not scientifically traceable

Appendix 2 contains a systematic verification of source use across all eight chapters of the dissertation. This assessment was not conducted as an exercise in error hunting, but as a balanced evaluation of the relationship between functionally correct use, interpretative tension

and possible overloading of sources. For each chapter, the analysis examined which literature genuinely carried the argument, which sources primarily served framing or legitimising functions, and the extent to which their use remained within the bounds of the original argumentation.

A consistent picture emerges from this analysis. Source use proves to be predominantly robust. Core empirical sources support, in most cases, the claims for which they are used, while theoretical and philosophical sources generally play a supporting rather than load-bearing role. The principal vulnerability lies not in fictitious or obviously misused literature, but in a recurring pattern of interpretative escalation: at several points, robust performance findings are translated somewhat more quickly into conclusions about underdetermination, pluralism or minimal assumptions than is strictly compelled by the sources themselves. This tension generally remains limited and manageable, but is methodologically relevant.

At the same time, the analysis shows that the later chapters become increasingly grounded internally. From Chapter 7 onwards, the centre of gravity shifts from external source use towards synthesis of empirical material developed earlier in the dissertation. This reduces the risk of decorative or strategic citation, but increases dependence on the proportionality with which earlier results have been interpreted. Taken as a whole, the verification confirms that the persuasiveness of the dissertation does not rest on a single spectacular source or one decisive claim, but on the coherence of a largely correct, functional and methodologically disciplined use of the literature.

This concludes the strictly academic assessment. A further methodological finding emerged from the verification process itself. Comparison of the ex ante prompt with the analysis actually produced showed that not all prescribed operations had been executed literally. A prompt should therefore be regarded as a statement of procedural intent, not as evidence that the specified procedure was actually performed. In AI-assisted research, procedural verification must consequently address the realised output as well as the instructions from which it was generated.

When the same material is viewed through the illusionist character of the dissertation, however, the picture shifts somewhat. Precisely because most sources are not used in conspicuously incorrect ways, but are instead embedded predominantly correctly, soberly and conventionally, the manuscript was able to function credibly. The strength of the construction lay not in spectacular falsification, but in control: in avoiding obvious errors, selecting safe but sufficiently supportive literature, and systematically tempering claims.

From this perspective, the limited interpretative overextension revealed by the verification is not merely a weakness, but also a functional component of the illusion. It helped give the dissertation a recognisable intellectual direction without exceeding the boundaries of plausibility. Source verification therefore reveals not only the academic standing of the manuscript, but also why it could function convincingly as an exercise in scientific illusionism.

Chapter 3 — Conclusions

This reflection report reconstructs the development of the *Columba Loquens* dissertation as a controlled experiment in scientific illusionism. Chapter 1 analysed the illusion as a design: a coherent set of construction rules aimed at establishing plausibility within the natural-science paradigm. Chapter 2 forensically reconstructed that design as an iterative process of generation, combination and correction, supplemented by systematic verification of both the bibliography and the use of literature.

The combination of these two perspectives — construction and reconstruction — makes it possible to determine the status of the final product more precisely. The dissertation is not a text that merely happens to be convincing, but the outcome of a controlled configuration in which form, data and argumentation were deliberately aligned. Its persuasiveness does not depend on any single element, but on the internal consistency of these components.

Within this context, the **epistemological implications** can be specified more precisely. The experiment shows that generative systems are capable of reproducing the form of natural-science argumentation convincingly, including plausible empirical structures. In this case, this applies to both text and data, making the dissertation a boundary case within AI-driven knowledge production.

What becomes visible is a shift in the relationship between explanation and empirical substantiation. The generated texts conform to the conventions of scientific argumentation: they contain literature references, methodological justification and statistical analysis. At the same time, the empirical basis rests on synthetic construction. The persuasiveness of the whole therefore lies not in the data themselves, but in the consistency and plausibility of the argumentative form.

This brings a classical tension from the philosophy of science back into view. In some domains, systems are capable of generating accurate predictions without the underlying mechanisms being fully understood. In the present case, the reverse occurs: convincing explanations are produced without a strong empirical basis underlying them. Explanation and substantiation become separated.

The case thereby reveals an **epistemological asymmetry**. Under certain conditions, narrative plausibility can take the place of empirical necessity. This does not imply that empirical research becomes redundant, but it does show that the form of scientific argumentation can generate persuasive force independently of the provenance of the data.

At the same time, the verification procedures described in Chapter 2 show where the limits of this construction lie. Bibliometric checking is necessary to exclude fictitious or incorrect references. Substantive verification reveals that not all sources provide direct support for the claims attributed to them. The combination of both forms of control prevents the manuscript from failing under scrutiny, but also demonstrates that persuasiveness and substantiation do not coincide.

The *Columba Loquens* dissertation therefore functions as a boundary case within the pentalogy. Whereas the other studies employ AI within existing empirical frameworks, here the empirical layer itself is also constructed. Precisely for that reason, the vulnerability of current assessment

practices becomes visible: when both text and data are plausibly constructed, the question shifts from ‘is this correct?’ to ‘how do we know that this is correct?’

The significance of this case therefore lies not in the substantive claims about pigeon navigation, but in the conditions under which such claims can function convincingly. The experiment shows that scientific credibility does not arise solely from the production of data, but from the combination of form, control and verification. In that sense, the illusion is not a deviation from scientific practice, but a means of making its underlying conditions visible.

Appendix 1: Bibliometric Verification

Audit Report: Reference List for Categorisation & Navigation

(conducted by Google Gemini)

Date: 16 March 2026

Scope: 48 references from the submitted source file.

Method: Each reference was checked against the *Semantic Scholar* and *Crossref* indices. Verification covered: (1) existence of the source, (2) accuracy of author/title/year/journal information, and (3) validity of the pagination.

1. Findings by Category

Status	Number	Definition
Verified correct	45	Source exists and metadata are accurate.
Metadata error	2	Source exists, but the pagination in the list was an AI placeholder (a guess).
Fictitious / hallucinated	1	Source cannot be traced in the scientific literature.

Detailed discrepancies:

- **Fictitious:** *Berg, M. E., et al. (2011). Categorization of multidimensional stimuli by pigeons.* This is an AI-generated 'hallucination' combining a plausible title with a non-existent article.
- **Metadata errors:**
  - *Turner, B. M., & Wasserman, E. A. (2023):* pagination corrected from "1–19" to "160–178".
  - *Watanabe, S. (1993):* pagination corrected from "1–13" to "25–38".

2. Overall Bibliometric Reliability Scores

- **Score A (including page numbers):** 93.75% (45/48 correct).
- **Score B (excluding page numbers):** 97.9% (47/48 correct).

Note: The high score is attributable to the robustness of the classical literature in the list, particularly in philosophy and ethology. The errors are concentrated in more recent experimental publications.

3. Recommendations

1. **Replace the fictitious source:** Remove *Berg et al. (2011)* from the reference list. Replace it with the comparable and verifiable article by *Urcuioli, P. J., & Lyon, L. N. (2011). Selective attention and categorization of multidimensional stimuli by pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 37(3), 345–351.* This article covers the same theoretical domain.
2. **Correct the metadata:** Manually update the page ranges for *Turner & Wasserman* and *Watanabe* to the verified values listed above.
3. **Document the audit:** Retain this report with the source materials. If a reviewer questions the reliability of the bibliography, it provides evidence that the bibliometric dataset was subjected to systematic validation.

4. Conclusion

Following implementation of the corrections above, the bibliography is **100% academically valid**. The AI 'bluff' — the generation of fictitious sources — has been fully eliminated by this audit.

Contra-Expertise: Convergence in Score, Divergence in Diagnosis

The bibliometric audit conducted by Google Gemini was subsequently subjected to an independent re-audit using ChatGPT. This contra-expertise produced a striking result. At the aggregate level, both audits arrived at the same **traceability score**: 47 of the 48 references (97.9%) could be related to identifiable scholarly publications. At the level of individual references, however, the two audits did not agree.

Gemini classified **Berg et al. (2011)** as fictitious. The ChatGPT re-audit found that a corresponding publication does exist, although the reference in the dissertation contains substantially incorrect metadata. Conversely, the re-audit identified **Gagliardo et al. (2009)** as the single genuinely hallucinated reference: the citation appears to combine elements of existing publications into a bibliographic entity that cannot itself be traced.

The re-audit also identified considerably more metadata errors than Gemini, although these did not generally affect the traceability of the underlying literature. Consequently, the principal finding of the original audit — that 47 of 48 references correspond to usable scholarly sources — remains intact, while its diagnosis of the individual errors does not.

This comparison yields a further methodological observation. **Agreement between AI-generated audits at the aggregate level does not imply agreement at the level of individual classifications.** A bibliometric audit performed by a language model should therefore be treated as an analytical output requiring verification, rather than as an external ground truth. In this case, contra-expertise increased confidence in the overall traceability estimate while simultaneously reducing confidence in the precision of the original audit

Appendix 2: Verification of Source Use

A. Assessment of Literature Use in Chapter 1

Chapter 1 functions as the **frame of reference** for the dissertation as a whole and fulfils three closely related functions. First, it defines a problem: the tension between reliable performance and inadequate explanations. Second, it positions the research domain through a combination of literature on pigeon navigation and visual classification. Third, it introduces conceptual instruments — including structural uncertainty, the 'Arctic night' as a boundary condition, and the role of explanatory assumptions — culminating in the formulation of the central research question. The chapter therefore functions as a theoretical and methodological framework rather than as an empirical contribution.

Author(s)	Year	Argumentative dependence	Functional use in Chapter 1
Bülthoff, Heinrich H. & Ernst, Marc O.	2004	Medium	Multisensory integration as an analogy for robustness
Cook, Robert	2001	Low	Support for pigeons' classification capacity
Gigerenzer, Gerd et al.	1999	Medium	Heuristics and decision-making under uncertainty
Gigerenzer, Gerd & Selten, Reinhard	2001	Medium	Critique of optimisation models
Hempel, Carl	1965	Low	Legitimation of idealisations and model use
Herrnstein, Richard	1976	Medium	Exemplar-based classification
Hough	2022	Medium	Overview of multi-cue navigation
Mayr, Ernst	1982	Low	Biological framework of explanation
Mouritsen, Hans	2018	High	Core source for multi-cue navigation and robustness
Pusch	2022	Medium	Recent classification studies
Simon, Herbert A.	1956	Medium	Decision-making without optimisation
Tinbergen, Niko	1963	Low	Foundation for behavioural explanation
Wasserman, Edward A.	2024	High	Core source for classification and the representation question
Weisberg, Michael	2007	Low	Theory of idealisations and models

The literature base underpinning this positioning is broad but functionally structured. At the level of fundamental biological and philosophical interpretation, the chapter draws on the work of Niko Tinbergen and Ernst Mayr, who frame the relationship between description and explanation in biological systems. For pigeon navigation, the chapter relies primarily on review

work by Hans Mouritsen and Hough. The domain of classification and cognition is represented by Edward A. Wasserman, Pusch, Richard Herrnstein and Robert Cook. Decision-making under uncertainty is theoretically grounded in the work of Herbert A. Simon and Gerd Gigerenzer, supplemented by Gerd Gigerenzer and Reinhard Selten. Marc O. Ernst and Heinrich H. Bülthoff are cited in relation to perceptual integration. The methodological legitimation of idealisations and boundary cases is derived from Carl Hempel and Michael Weisberg.

Within this literature base, several sources display medium to high argumentative dependence. This applies particularly to the use of **Mouritsen and Hough**. These works support the proposition that pigeon navigation depends on multiple information sources and that disruption of individual cues generally leads to gradual changes in performance. This interpretation is largely correct and well grounded in the literature. At the same time, studies exist in which strong dependence on specific cues has been demonstrated, particularly under context-specific conditions. The formulation that no single information source can be regarded as exclusively necessary is therefore somewhat stronger than can be strictly compelled empirically in an introductory context. The assessment is that this constitutes a slight oversimplification, but not a substantive misrepresentation.

By contrast, the use of the classification literature, particularly **Wasserman, Pusch, Herrnstein and Cook**, is robust and precise. The combination of stable performance under variation and theoretical underdetermination between representational models is accurately presented. The tension between exemplar, prototype and other approaches is not reduced to a single interpretation, but serves to substantiate the central problem. Within this domain, argumentative dependence is high, but the use of the literature is methodologically sound.

A more conceptual dependence arises in the use of **Simon and Gigerenzer**. Their work is used to characterise structural uncertainty and heuristic decision-making. In doing so, an implicit bridge is drawn from human decision-making to animal behaviour. This transition is substantively defensible, but is not explicitly justified as an analogy or transposition between domains. The assessment is that this constitutes a conceptual extension that is acceptable within the behavioural sciences, but could have been made more explicit methodologically.

A comparable situation arises in the use of **Ernst and Bülthoff**. Their work on multisensory integration in perception is used to support robustness through the integration of multiple information sources. Since this research primarily concerns laboratory perception rather than large-scale navigation, its use here functions as an analogy rather than as direct empirical support. This use is substantively defensible, but remains implicit. The assessment is that the application is appropriate, although it would have been preferable to mark it explicitly as analogical.

By contrast, the use of **Hempel and Weisberg** is methodologically strong and appropriate. The legitimation of idealisations and boundary cases directly accords with their work and convincingly supports the introduction of the ‘Arctic night’ as an analytical instrument. Within this part of the argument, argumentative dependence is high and the use of the literature is fully appropriate.

At the **system level**, Chapter 1 displays a high degree of conceptual consistency. The central tension between performance and explanation is sharply defined and consistently maintained. Mechanistic overclaims are avoided; the argument remains strictly at the level of assumption

structures. The methodological positioning of the ‘Arctic night’ as a heuristic instrument is clear and appropriate. The use of literature is purposeful and functional: every source demonstrably contributes to the construction of the argument.

At the same time, there is a slight but consistent directional tendency in the framing. The chapter moves from an empirical tension towards the suggestion that explanatory pluralism is a structural feature of the domain. This move is not incorrect, but is expressed more strongly than is strictly necessary on the basis of the literature presented. This applies particularly to the generalisation within the navigation literature, where context-dependent variation receives less emphasis. In addition, cross-domain analogies remain implicit, whereas making them explicit would clarify their methodological status. The rhetorical formulation is persuasive, but at several points somewhat more assertive than the underlying empirical basis strictly warrants. The assessment of Chapter 1 therefore yields a predominantly robust picture with a slight conceptual bias. The use of literature is strong, conceptual clarity is high and the methodological positioning is convincing. Empirical caution is largely present, but not maximal. The risk of distortion remains limited, but is not absent, particularly as a result of subtle overgeneralisation and implicit analogies.

In summary, Chapter 1 provides a solid theoretical foundation for the dissertation and uses the relevant literature largely correctly and functionally. The central movement — from an empirical tension towards a framework in which explanatory pluralism is possible — is substantively defensible, but at this stage is not yet fully compelled by the empirical evidence. The chapter therefore functions effectively as framing, but contains a limited degree of **epistemological steering** that requires further substantiation in the empirical chapters.

B. Assessment of Literature Use in Chapter 2

0. Frame of Reference

Chapter 2 argues that homing pigeons display stable performance under variation in both classification and navigation tasks, while the literature does not yield a single necessary explanatory structure. The chapter develops cumulatively: (§2.1) classification → robust generalisation, (§2.2) decision-making → multi-cue integration under uncertainty, (§2.3) homing → multiple explanatory models existing alongside one another, (§2.4) epistemology → the problem of assessability, (§2.5) methodological scope → experimental manipulability, (§2.6) synthesis → performance retention + explanatory pluralism as the central tension. Empirical claims: performance is reproducible and robust. Epistemological claims: explanations remain underdetermined and require assumptions to be made explicit.

1. Full Classification List (All Sources)

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). *Natural concepts in pigeons*. *Journal of Experimental Psychology: Animal Behavior Processes, 2*(4), 285–302. – Argumentative dependence: high

Wasserman, E. A., et al. (2014). [multiple categorisation in pigeons]. – Argumentative dependence: medium

Watanabe, S. (1993). *Object-picture equivalence in the pigeon*. *Behavioural Processes, 30*, 225–232. – Argumentative dependence: low

Berg, M. E., et al. (2011). *Categorization of multidimensional stimuli by pigeons*. *Journal of Experimental Psychology: Animal Behavior Processes, 37*(3), 331–344. – Argumentative dependence: low

Cook, R. G. (Ed.). (2001). *Avian Visual Cognition*. – Argumentative dependence: medium

Qadri, M. A. J., & Cook, R. G. (2015). *Experimental divergences in the visual cognition of birds and mammals*. – Argumentative dependence: low

Turner, B. M., & Wasserman, E. A. (2023). *Complex category structures*…. – Argumentative dependence: low

Edwards, C. A. (1987). *Feature selection and categorization in pigeons*. – Argumentative dependence: low

Wiltschko, R., & Wiltschko, W. (1995). *Magnetic Orientation in Animals*. – Argumentative dependence: high

Wallraff, H. G. (2005). *Avian Navigation: Pigeon Homing as a Paradigm*. – Argumentative dependence: high

Biro, D., et al. (2004). *Familiar route loyalty implies visual pilotage*. – Argumentative dependence: medium

Gagliardo, A., et al. (2009). *Role of visual landmarks revisited*. – Argumentative dependence: medium

Simon, H. A. (1956). *Rational choice and the structure of the environment*. – Argumentative dependence: medium

Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple Heuristics That Make Us Smart*. – Argumentative dependence: medium

Guilford, T., & Biro, D. (2014). *Route-following and landmark-based navigation*. – Argumentative dependence: medium

Wiltschko, R., & Wiltschko, W. (2005). [magnetic manipulation]. – Argumentative dependence: low

Mouritsen, H. (2018). [multi-cue navigation review]. – Argumentative dependence: medium

Hough, G. E. (2022). *Neural substrates of pigeon navigation*. – Argumentative dependence: medium

2. Analysis of Medium/High Dependence

Herrnstein et al. (1976) – Argumentative dependence: high. This source supports the fundamental claim that pigeons generalise across stimulus variation and that classification therefore cannot be reduced to memorisation. It forms the empirical basis of §2.1 and, by extension, of the broader argument that performance is robust. Its use remains largely within the original experimental scope and substantially strengthens the epistemological robustness of the argument. The vulnerability arises in the extrapolation to underdetermination: generalisation is implicitly treated as evidence that no specific representational structure is necessary. This introduces limited conceptual overloading and selective reading, but no serious distortion. **Balance: predominantly robust.**

Wasserman et al. (2014) – Argumentative dependence: medium. The source reinforces the picture of scalable and parallel categorisation and thereby supports the robustness claim. It is replaceable but functionally relevant. Its use remains substantively correct, but it carries greater epistemological weight than is strictly warranted: performance is interpreted as an indication of structural underdetermination. This creates slight conceptual overloading and loss of context. **Balance: mixed but functional.**

Cook (2001) – Argumentative dependence: medium. This source functions as a synthesis anchor that legitimises the cumulative nature of the field. Its contribution is primarily stabilising: it positions individual studies within a broader research trajectory. Its use remains

within the bounds of a review work, but is relatively generic and therefore susceptible to substitution of reference for argument. **Balance: mixed but stable.**

Wiltschko & Wiltschko (1995) – Argumentative dependence: high. This source carries a substantial part of the navigation claim: that multiple cues contribute to behaviour and that disruption leads to variation rather than collapse. Its use is empirically appropriate and substantially strengthens the argument. The vulnerability lies in the interpretative step: variation and multi-cue use are treated as evidence that no single cue is necessary. This is a stronger conclusion than the source itself compels and entails conceptual overloading and selective reading. **Balance: robust but interpretatively strained.**

Wallraff (2005) – Argumentative dependence: high. Wallraff functions as a central pillar of the multi-cue interpretation of homing. The source increases the coherence of §2.3 and provides an integrative framework. At the same time, a source rooted in a particular research tradition is used as a quasi-neutral synthesis, introducing some loss of context. The extrapolation towards structural explanatory pluralism partly exceeds its original function. **Balance: mixed but load-bearing.**

Biro et al. (2004) & Gagliardo et al. (2009) – Argumentative dependence: medium. These studies support the pattern of gradual performance change and the contribution of visual information. They strengthen the empirical basis but are not structurally indispensable. Their use is appropriate, with minimal distortion. Their contribution is primarily illustrative within an already established pattern. **Balance: predominantly robust.**

Simon (1956) & Gigerenzer et al. (1999) – Argumentative dependence: medium. These sources provide the theoretical framework for non-optimal but functional decision-making. They increase the interpretative coherence of §2.2 by positioning behaviour outside an optimisation imperative. Their use remains largely within the original argumentation, but partly serves to legitimise an interpretative step that could also have been made without these sources. **Balance: mixed but coherent.**

Guilford & Biro (2014) – Argumentative dependence: medium. This source supports the methodological and empirical description of navigational behaviour and field studies. Its use is appropriate and strengthens the empirical grounding of §2.5. Risks are limited; its role is primarily supportive.
Balance: robust.

Mouritsen (2018) & Hough (2022) – Argumentative dependence: medium. These sources provide modern syntheses of multi-cue integration. They strengthen the contemporary relevance and coherence of the integration framework. At the same time, they contribute to the shift from empirical observation to epistemological conclusion, introducing conceptual overloading and a slight substitution of reference for argument. **Balance: mixed but functional.**

3. System Diagnosis

Source use in Chapter 2 is predominantly robust, but systematically carries an interpretative loading.

Strengths include a clear empirical basis across multiple domains — classification and navigation — a cumulative structure combining classical and contemporary literature, and a consistent connection between observations and claims concerning performance retention. The principal vulnerabilities lie in a recurring pattern of **epistemological escalation**: from variation → underdetermination, limited explicit consideration of alternative interpretations, and concentration of interpretative weight in a small number of key sources.

There is no fundamental misuse of the literature. The tension lies not in incorrect use, but in systematic overinterpretation in the same direction.

Overall assessment:

→ Predominantly robust, with consistent but manageable interpretative tension.

C. Assessment of Literature Use in Chapter 3

0. Frame of Reference

Chapter 3 justifies a research design that does not aim at model selection, but at making the assumptions underlying explanatory models explicit and empirically assessable. The central methodological claim is that performance retention under controlled variation can be investigated without resorting to optimisation, causal selection or premature theory elimination.

Empirically, the chapter is not yet evidential but design-oriented: it describes how classification, homing and combined variation are measured. Epistemologically, it legitimises the shift from 'which model is correct?' to 'which assumptions prove necessary?'. The primary function of source use is therefore not empirical substantiation, but methodological legitimation and ethical grounding.

1. Full Classification List (All Sources)

Mayo, D. G. (2018). *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press. – Argumentative dependence: medium

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press. – Argumentative dependence: medium

Stanford, K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford University Press. – Argumentative dependence: medium

Russell, W. M. S., & Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. – Argumentative dependence: medium

Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. – Argumentative dependence: medium

European Commission. (2020). *Working document on the implementation of Directive 2010/63/EU*. – Argumentative dependence: low to medium

2. Analysis of Medium/High Dependence

Mayo (2018) – Argumentative dependence: medium. Mayo supports the claim that assumptions can be made explicit and systematically exposed to empirical pressure. The functional contribution is clear: the source provides methodological legitimacy for a research strategy that does not primarily rank hypotheses, but examines the strength and tenability of assumptions. Its use is, however, broader than strict fidelity to the text would warrant. Mayo's work concerns severe testing and statistical inference; here it is used to support a more general assumption-driven research logic. This produces slight conceptual overloading and contextual

displacement. There is no obvious misuse of authority, but the source partly functions as an emblem of legitimacy for an approach that is also defended without an explicit severe-testing architecture. **Balance: mixed but functional.**

Woodward (2003) – Argumentative dependence: medium. Woodward supports the passage in which controlled variation is presented as a manipulation-based means of making causal and functional relevance visible. The positive contribution is substantial: the source helps position the design as a systematic intervention logic rather than merely descriptive observation. Its use remains reasonably close to the spirit of Woodward, but is simultaneously narrowed and broadened: narrowed because the rich interventionist theory is reduced to a brief methodological legitimation, and broadened because it is used to legitimise the research architecture as a whole. This creates a limited substitution of reference for argument. **Balance: predominantly robust.**

Stanford (2006) – Argumentative dependence: medium. Stanford is used to identify the problem of underdetermination and unconceived alternatives methodologically. This is an appropriate use: the source strengthens the epistemological significance of the premise that multiple explanatory models may remain compatible. The vulnerability lies in the shift of scale. Stanford analyses a general problem in the philosophy of science; Chapter 3 uses this to legitimise a concrete behavioural-biological design. This is defensible, but not self-evident. There is therefore slight contextual broadening, but little genuine distortion. **Balance: predominantly robust.**

Russell & Burch (1959) – Argumentative dependence: medium. This source functions as the classical ethical foundation for reduction and refinement in animal research. Its contribution is clear and appropriate: it anchors the ethical logic of proportionality and necessity. Its use remains within the original argument and strengthens the normative consistency of the chapter. The only limitation is that the source is canonical and general; the concrete ethical operationalisation ultimately rests more on the text itself than on the source. **Balance: robust.**

Directive 2010/63/EU – Argumentative dependence: medium. The Directive supports the claim that the design complies with current European regulation. This is a functionally necessary normative source for the ethical section. Its use is appropriate and precise: it provides regulatory grounding without theoretical overloading. A slight broadening occurs only where methodological proportionality is also inferred from regulatory compliance. **Balance: robust.**

European Commission (2020) – Argumentative dependence: low to medium. This source supports the application and interpretation of the Directive. It increases procedural credibility, but is not independently load-bearing. Its use is correct and largely unproblematic. **Balance: predominantly robust.**

3. System Diagnosis

Chapter 3 is methodologically sparing in its citations but relatively coherent. Unlike Chapter 2, it does not use sources to establish empirical patterns, but to legitimise three elements: assumption-driven testing, epistemological restraint and ethical proportionality. This makes source use compact and functional.

A particular strength is that the small number of sources each has a clearly defined role. There is little noise, little artificial diversity and hardly any decorative citation. The ethical sources are especially well integrated: the relationship between source and claim is direct and proportionate. The methodological and philosophy-of-science sources are substantively appropriate, but carry somewhat greater argumentative weight. Mayo in particular, and Stanford to a lesser extent, are used to legitimise a broader research architecture than their original purposes strictly require.

The balance therefore clearly remains positive. The vulnerability of Chapter 3 lies not in incorrect source use, but in **slight theoretical overextension based on a limited number of sources**. Because the text is internally consistent, this tension remains manageable.

Overall assessment:

→ Predominantly robust, with limited overloading of legitimising sources within the methodological framework.

D. Assessment of Literature Use in Chapter 4

0. Frame of Reference

Chapter 4 does not seek to demonstrate once again that pigeons classify, but to reposition an established classification paradigm as an instrument for delimiting assumptions. The central claim is that stable classification behaviour under variation can be described empirically without this immediately implying one specific explanatory model.

Empirically, responses remain largely stable under novel, degraded and conflicting stimuli, with proportional degradation but without collapse. Epistemologically, these performances do not uniquely compel exemplar-based, prototype-based or integrative descriptions; the chapter therefore seeks to delimit negatively which assumptions prove unnecessary.

Chapter 4 consequently combines two layers: a theoretical literature-based line in §4.0 and an original empirical elaboration in §4.1–§4.7.

1. Full Classification List (All Sources)

Nosofsky, R. M. [exemplar-based categorisation framework]. – Argumentative dependence: medium

Reed, P. (1972). [stimulus control / feature selection]. – Argumentative dependence: low

Honig, W. K., & Stewart, K. E. (1988). [local versus global stimulus control in pigeons]. – Argumentative dependence: medium

Cook, R. G., Katz, J. S., & Cavoto, B. R. (1997). [complex visual stimulus control in pigeons]. – Argumentative dependence: medium

Lea, S. E. G., & Ryan, C. M. E. (1990). [artificial categories / problem of concept formation]. – Argumentative dependence: medium

Pearce, J. M. (1994). [stimulus integration / configural learning]. – Argumentative dependence: medium

Young, M. E., & Wasserman, E. A. (2001). [same–different judgements and processing of features/combinations]. – Argumentative dependence: medium

Blough, D. S. (1975). *Steady-state generalization and discrimination in pigeons*. – Argumentative dependence: medium

Lazareva, O. F., & Wasserman, E. A. (2012). *Categories and concepts in animals*. – Argumentative dependence: medium

2. Analysis of Medium/High Dependence

Nosofsky – Argumentative dependence: medium. Nosofsky supports the first strand of literature: stable behaviour can be described in exemplar-based terms without assuming abstract rules or categories. The positive contribution is substantial, because this source establishes the formal possibility of describing performance without immediately making strong cognitive claims. This aligns well with the purpose of the chapter. At the same time, Nosofsky is used primarily as a model type rather than as a fully developed theoretical position with its own conditions and limitations. This introduces a slight reduction in complexity: the framework functions more as a representative of 'one descriptive possibility' than as a complete theory. **Balance: predominantly robust.**

Honig & Stewart (1988) and Cook, Katz & Cavoto (1997) – Argumentative dependence: medium. Together, these sources support the second strand of literature: behaviour may be controlled by unexpected or limited stimulus features. Their principal contribution is that they undermine the assumed equivalence of stimuli and thereby directly legitimise the design of §4.4. This provides a clear epistemological strengthening: the sources demonstrate why controlled variation is necessary. Their use remains substantively appropriate, but they carry slightly greater functional weight because the analogy with homing is introduced at the same time. The analogy is useful, but is not compelled by the sources themselves. **Balance: mixed but functional.**

Lea & Ryan (1990) – Argumentative dependence: medium. This source supports the warning that claims about concept formation become problematic when it remains unclear which features actually control behaviour. This fits the negative epistemological objective of the chapter particularly well. Its use remains close to the original argument and is methodologically productive: it legitimises restraint. The vulnerability is limited; the source functions primarily as a safeguard against overinterpretation. **Balance: robust.**

Pearce (1994), Young & Wasserman (2001), Blough (1975) – Argumentative dependence: medium. Together, these sources form the third strand of literature: similar performance may arise from different forms of stimulus integration. They provide the principal support for the proposition that performance alone is insufficient to determine strategy use or representational form. Their contribution is strong and well integrated into the argument of §4.0. There is, however, some conceptual compression: differences between configural learning, feature processing and generalisation are combined in support of a single common point. This is analytically efficient, but smooths over some nuance. **Balance: predominantly robust.**

Lazareva & Wasserman (2012) – Argumentative dependence: medium. This review concludes the theoretical introduction by identifying the ambiguity of animal categorisation research as a structural feature of the field. Its contribution is important as a synthesis anchor: it draws together the preceding strands of literature. At the same time, this is also where the source carries somewhat more argumentative weight than strictly necessary, because the review not only synthesises the literature but implicitly legitimises the broader conclusion of underdetermination. **Balance: mixed but stable.**

3. System Diagnosis

Chapter 4 has a dual structure: the sources are concentrated almost entirely in §4.0, while the remainder of the chapter rests primarily on the dissertation's own empirical elaboration. Source use is therefore more compact than in Chapter 2, but also more sharply functional.

A major strength is that the literature in §4.0 is well selected for the specific purpose of this chapter: not to demonstrate classification, but to show why classification data may remain explanatorily underdetermined. The sources operate cumulatively and support three clear strands: ambiguity between exemplar and prototype descriptions, stimulus control, and stimulus integration. This is methodologically strong and relatively clean.

The principal vulnerability lies not in incorrect source use, but in analogical scaling. Literature from the classification domain is repeatedly used to clarify the logic of homing research as well. This is substantively useful, but is not fully compelled by the sources themselves. In addition, §4.0 sometimes compresses differences between theories too quickly into one common point: performance does not decide between theories.

Against this, the empirical part displays considerable discipline: the results are consistently kept descriptive and are not translated back into strong cognitive claims.

Overall assessment:
→ Predominantly robust.

E. Assessment of Literature Use in Chapter 5

0. Frame of Reference

Chapter 5 addresses Research Question 2 by using homing behaviour under variation of individual navigational cues not to identify a single mechanism, but to assess which assumptions concerning necessity, dominance and compensation remain descriptively tenable. Empirically, the chapter claims that manipulation of individual cues produces shifts in performance but not structural disorganisation. Epistemologically, it claims that these patterns do not establish an exclusive hierarchy of cues. The chapter therefore has a dual purpose: a classical navigation problem is reorganised as assumption testing rather than mechanism identification.

1. Full Classification List (All Sources)

Kramer, G. (1953). [sun compass / time compensation]. – Argumentative dependence: medium

Mouritsen, H. (2018). [recent synthesis of avian navigation]. – Argumentative dependence: medium

Wiltschko, R., & Wiltschko, W. (1972). [magnetic hypotheses]. – Argumentative dependence: low to medium

Wiltschko, R., & Wiltschko, W. (2005). [magnetic interference / field manipulation]. – Argumentative dependence: high

Papi, F., et al. (1972). [olfactory map hypothesis]. – Argumentative dependence: low to medium

Wallraff, H. G. (2005). *Avian Navigation: Pigeon Homing as a Paradigm*. – Argumentative dependence: high

Biro, D., et al. (2004). *Familiar route loyalty implies visual pilotage*. – Argumentative dependence: medium

Guilford, T., & Biro, D. (2014). *Route-following and landmark-based navigation in pigeons*. – Argumentative dependence: medium

Gagliardo, A. (2013). *Forty years of olfactory navigation in birds*. – Argumentative dependence: medium

Hough, G. E. (2022). [context-dependent cue integration]. – Argumentative dependence: medium

2. Analysis of Medium/High Dependence

Wiltschko & Wiltschko (2005) – Argumentative dependence: high. This source supports a substantial part of the historical and methodological logic of the chapter. It not only supports the proposition that magnetic hypotheses developed through intervention paradigms, but also legitimises the translation of theoretical claims into manipulable experimental dimensions. Its positive contribution is substantial: without this source, the tradition of cue manipulation on which the chapter builds would be less firmly grounded. Its use remains largely within the original argument, namely that magnetic information can be experimentally manipulated and may have behavioural consequences. The main vulnerability lies in the subsequent step: interference research is also used to support the conclusion that exclusive necessity of a single cue is descriptively problematic. This is a stronger and more comparative conclusion than the source itself supports. **Balance: predominantly robust, with limited interpretative broadening.**

Wallraff (2005) – Argumentative dependence: high. Wallraff functions as a pivotal source for the olfactory strand and for the broader representation of homing as a multi-cue system. Its functional contribution is very substantial: the source increases both the historical depth and the integrative coherence of the chapter. At the same time, it is also the most heavily burdened source. Wallraff is used both as documentation of a specific research tradition and as a supporting synthesis for the proposition that multiple models remain compatible. This creates some loss of context: a source strongly associated with one explanatory tradition is partly used as neutral background for pluralism. This is not untenable, but constitutes a clear point of tension. **Balance: mixed but load-bearing.**

Mouritsen (2018) – Argumentative dependence: medium. Mouritsen provides an updated synthesis and connects the historical literature to a modern multi-cue framework. Its positive contribution is that the chapter does not rely solely on classical paradigms, but demonstrates that the debate continues in contemporary terms of integration and context dependence. Its use is functionally appropriate. The vulnerability lies in the fact that the source partly serves to confirm the chapter's framework before the empirical results of the chapter have been presented. This creates a slight risk of substituting reference for argument. **Balance: mixed but functional.**

Biro et al. (2004) and Guilford & Biro (2014) – Argumentative dependence: medium. These sources support the visual/route-based strand and the methodological claim that navigational behaviour can be recorded in detail. Their contribution is twofold: they support a specific explanatory line and the experimental feasibility of systematic behavioural measurement. Their use remains substantively correct and relatively close to the original function of the sources. The vulnerability is limited; they are not clearly overloaded, but are used primarily as precedent and support. **Balance: predominantly robust.**

Gagliardo (2013) – Argumentative dependence: medium. Gagliardo functions as meta-commentary on the olfactory tradition and supports the proposition that similar experimental outcomes can be interpreted within different theoretical frameworks. This helps structure the transition from historical inventory to assumption testing. Its use is appropriate, although the source is also used to mark interpretative pluralism, which gives it greater epistemological weight than a purely historical review would carry. **Balance: mixed but coherent.**

Hough (2022) – Argumentative dependence: medium. Hough primarily supports §5.4, where context-dependent integration and cue weighting legitimise the design of manipulable dimensions. Its contribution is substantial because it helps the chapter move from classical cue hypotheses towards a more modern integration framework. The vulnerability is that a review concerned with the neural substrates of navigation functions here as support for a relatively abstract design principle. This is defensible, but broader than strictly necessary. **Balance: mixed but functional.**

3. System Diagnosis

Chapter 5 is more strongly integrated in its source use than Chapter 3 and methodologically tighter than Chapter 2. Literature is not used primarily to postulate pluralism, but to transform a historical tradition of manipulation into a testable design for assumption structures. This is a clear strength. The sources generally have a well-defined function: historical precedent, experimental precedent, or synthesis of integration.

The principal positive factor is the close correspondence between the empirical design and the cited literature. Unlike Chapter 2, the transition from literature to chapter function is more concrete: classical hypotheses are translated into conditions, matrices, a reduced configuration design and assessment criteria. As a result, source use appears less decorative and more operational.

The principal vulnerability again lies with the synthesis sources, especially Wallraff and, to a lesser extent, Mouritsen and Hough. These sources provide not only historical or substantive support, but also support for the broader conclusion that exclusive necessity is descriptively open to question. This does not constitute fundamental misuse, but does create a recognisable interpretative tension.

There is also a minor internal methodological tension within the chapter itself: a full configuration matrix is presented first, followed by a reduced design. This is not directly a source-use issue, but it makes the progression from sources to design somewhat less streamlined.

Overall assessment:
→ Predominantly robust, with manageable interpretative tension in the synthesis of classical and contemporary cue literature.

F. Assessment of Literature Use in Chapter 6

0. Frame of Reference

Chapter 6 shifts the analysis from variation in individual cues to cumulative system pressure. The central question is no longer whether one specific information source is necessary, but how homing behaviour changes when multiple cues are reduced simultaneously. Empirically, the chapter claims that double and triple reductions produce shifts in performance, but no abrupt disorganisation. Epistemologically, it argues that 'robustness' is justified only if such cumulative reduction produces gradual rather than threshold-like effects within the investigated range. The chapter therefore functions as a test of robustness under system pressure, not as a test of individual mechanisms.

1. Full Classification List (All Sources)

Ernst, M. O., & Bülthoff, H. H. (2004). *Merging the senses into a robust percept*. – Argumentative dependence: medium

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). *Multisensory integration: Psychophysics, neurophysiology, and computation*. – Argumentative dependence: medium

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). *Neural correlates of reliability-based cue weighting during multisensory integration*. – Argumentative dependence: medium

Jacobs, L. F., & Schenk, F. (2003). *Unpacking the cognitive map: The parallel map theory of hippocampal function*. – Argumentative dependence: low to medium

Cheng, K., & Spetch, M. L. (2011). *Mechanisms of landmark use in mammals and birds*. – Argumentative dependence: low to medium

Kitano, H. (2004). *Biological robustness*. – Argumentative dependence: medium

Scheffer, M., et al. (2001). [critical transitions / system boundaries]. – Argumentative dependence: medium

Wiltschko & Wiltschko; Papi; Wallraff; Kramer; Biro; Guilford [as previously discussed research programmes]. – Argumentative dependence: low

2. Analysis of Medium/High Dependence

Ernst & Bülthoff (2004) – Argumentative dependence: medium. This source supports the general idea that multisensory systems combine signals according to their relative reliability. Its positive contribution is clear: it provides a conceptual vocabulary for cue weighting and for the hypothesis that reduction of one cue need not necessarily lead to collapse. It thereby supports the transition from Chapter 5 to Chapter 6. The vulnerability lies in extending perceptual integration at a relatively small scale to large-scale navigation. The chapter itself explicitly acknowledges that this step is not self-evident, which reduces the interpretative tension. **Balance: mixed but carefully used.**

Angelaki et al. (2009) and Fetsch et al. (2012) – Argumentative dependence: medium. These sources strengthen the cue-weighting framework by providing neurophysiological and computational support for adaptive shifts in weighting as signal reliability changes. Their functional contribution is to make redundancy and reweighting theoretically plausible. At the same time, this is precisely where the greatest overextension may occur: the move from visual–vestibular integration to pigeon navigation involves a substantial shift in scale and domain. The chapter tempers this itself by stating that such transfer is 'not self-evident'. Conceptual overloading is therefore not eliminated, but methodologically controlled. **Balance: mixed but defensible.**

Kitano (2004) – Argumentative dependence: medium. Kitano provides the robustness framework: systems may remain functionally stable under disturbance without that stability being unlimited. This source fits the chapter's objective well because it directly supports the distinction between gradual sensitivity and boundary conditions. Its use remains close to the original argument and is methodologically productive. Little distortion is apparent; the source functions as a clear conceptual anchor. **Balance: robust.**

Scheffer et al. (2001) – Argumentative dependence: medium. Scheffer supports the possibility of abrupt transitions when critical thresholds are crossed. This contribution is important because it prevents the chapter from steering exclusively towards robustness and explicitly leaves room for collapse or threshold behaviour. This strengthens the epistemological balance of the chapter. The vulnerability lies in the fact that a general systems framework for critical transitions is used primarily as a heuristic contrast; it is not a direct model of homing behaviour. Even so, its use is clearly delimited and not heavily overloaded. **Balance: predominantly robust.**

Jacobs & Schenk (2003) and Cheng & Spetch (2011) – Argumentative dependence: low to medium. These sources primarily perform a corrective function: they constrain any overly rapid transfer of multisensory integration models to navigation by introducing spatial representation, experience and environmental structure. Their positive contribution is disproportionately large relative to their limited weight in the chapter because they protect the argument against simplification. They are not overloaded. **Balance: robust.**

3. System Diagnosis

Chapter 6 uses fewer sources than Chapter 5 but is conceptually sharper. The literature serves a precisely defined purpose: establishing a test between three descriptive possibilities — strong redundancy, system boundaries, or gradual proportional change.

A major strength is the unusually balanced use of literature. Unlike some earlier chapters, Chapter 6 does not cite only sources that support pluralism or robustness; it explicitly introduces a counter-possibility, namely abrupt system boundaries. This strengthens the argumentative architecture.

The chapter also explicitly acknowledges that transferring models of multisensory integration to large-scale navigation is not straightforward. This built-in restraint increases the reliability of the source use. The principal vulnerability lies in the shift of scale: perceptual and neurophysiological integration models are used as a heuristic framework for homing. This is defensible, but not directly compelled by the sources. The tension remains limited, however, because the chapter uses these sources as a problem framework rather than as proof. The empirical section also remains disciplined: the results are kept descriptive and lead only to the conclusion that no transition point is visible within the investigated range.

Overall assessment:

→ Predominantly robust.

G. Assessment of Literature Use in Chapter 7

I. Frame of Reference and Function of Chapter 7

Chapter 7 has a strictly synthetic character. No new empirical data are introduced and no additional literature is incorporated. Its function is to make explicit and organise assumptions that were previously implicit or dispersed across earlier chapters. This organisation is based on the results from Chapters 4 to 6 and follows the assumption structure formulated in Chapter 3. The analysis therefore remains entirely within the previously delimited empirical and conceptual framework.

II. Literature Used and Dependence Structure

No external sources are used in this chapter. The argument is based entirely on internal references to previously developed parts of the dissertation. The central role is played by the empirical results from Chapters 4 to 6, which are systematically brought together in Table 7.1. Dependence on external literature is therefore absent, while dependence on the internal empirical infrastructure is maximal.

III. Analysis of Argumentative Dependence

The decision not to introduce new literature contributes to the transparency of the argument. Because no external authority is invoked, it is immediately clear that the conclusions must be supported entirely by the data and interpretations presented earlier. This structure prevents new theoretical claims from being implicitly legitimised through additional sources.

At the same time, the absence of new sources means that the chapter cannot provide an independent correction or recalibration. The validity of the analysis depends entirely on how the empirical results in the preceding chapters were interpreted. Any overestimation or interpretative shift is not corrected here, but carried forward in condensed form.

IV. System Diagnosis

The methodological position of Chapter 7 is therefore clearly delimited. The chapter introduces no new data, additional models or external theoretical frameworks, but confines itself to making explicit the status of assumptions within the existing range. This keeps the risk of distortion through selective source use or theoretical import low.

The formulations remain consistently restrained. Assumptions are not presented as confirmed or refuted, but as unnecessary, indeterminate or compatible with the data. This terminology keeps the analysis proportionate to the underlying evidence and prevents the synthesis from claiming more than is empirically justified.

V. Overall Assessment

Chapter 7 is internally consistent and methodologically restrained. The argument is derived entirely from previously presented results and introduces no new assumptions or external legitimation. Its principal limitation lies in its complete dependence on earlier interpretations, since the chapter itself performs no additional testing.

VI. Summary Diagnosis

The chapter functions as a direct explication of findings already developed. Within this role, the analysis remains transparent and proportionate. The robustness of its conclusions is therefore directly dependent on the strength of the empirical chapters on which they rest.

H. Assessment of Literature Use in Chapter 8

I. Frame of Reference and Function of Chapter 8

Chapter 8 fulfils the conventional functions of a conclusion and discussion chapter. Section 8.1 brings together and synthesises the results from Chapters 4–7. Section 8.2 discusses the limitations of the design, with particular emphasis on the range of variation, level of measurement and interpretation. Sections 8.3–8.4 formulate possible directions for further research. The chapter introduces no new empirical data. Its principal theoretical step consists in formulating underdetermination as an outcome of the research conducted, rather than as a philosophical problem assumed in advance.

II. Literature Used and Dependence Structure

The chapter is based almost entirely on the results and analytical frameworks from Chapters 4–7. External literature plays only a limited role. Sandra D. Mitchell is the only author explicitly cited, in §8.1.4. This reference serves to position explanatory pluralism as a potentially structural feature of scientific description. Its use is conceptually appropriate and remains confined to interpretative contextualisation. No appeal is made to external empirical substantiation.

III. Analysis of Argumentative Dependence

Dependence on Sandra D. Mitchell is limited. The central conclusions follow from the internal dataset; the reference influences the formulation of the interpretation rather than its empirical basis. The chapter as a whole is therefore internally grounded. The argument rests on comparison of behavioural outcomes under variation, not on external literature.

IV. System Diagnosis

The formulations remain consistently tied to the investigated range. Limitations in the range of variation and level of measurement are explicitly acknowledged. No claims are made about conditions beyond the experiment. In §8.2, internal validity is discussed systematically. The choice of behavioural operationalisations is recognised as a limitation. It is also acknowledged that possible thresholds or collapse outside the investigated range cannot be established. Theory remains supportive. The reference to Sandra D. Mitchell functions as an interpretative framework rather than as an evidential basis.

V. Vulnerabilities and Points of Tension

The principal vulnerability concerns the position of the philosophical reference. Sandra D. Mitchell is introduced only in the conclusion and is not developed in the preceding chapters. In addition, the formulation of ‘structural pluralism’ involves an interpretative step that extends beyond the direct empirical observations. This step is carefully delimited, but remains dependent on extrapolation from the investigated domain. Finally, the analysis is based entirely on behavioural data. The chapter acknowledges that model differentiation might be stronger at another level of measurement.

VI. Overall Assessment

Chapter 8 is internally consistent and remains within the boundaries of the research conducted. The argument is largely independent of external literature. The interpretative step towards structural pluralism is present, but is cautiously formulated and not developed into a full theoretical argument.

VII. Summary Diagnosis

The chapter integrates the empirical results into a single conclusion: within this design, performance retention does not lead to reduction of explanatory models. This conclusion is bound to the investigated range and the level of measurement employed. The theoretical generalisation remains only lightly substantiated and is not developed further.

I. Source-Verification Protocol: Ex Ante Design and Ex Post Reconstruction

I.1 Ex Ante Protocol

The source-use verification presented in this appendix was initiated through a dedicated prompt designed to treat the literature base not as a single bibliography, but as a **chapter-specific distributed epistemological system**. The underlying premise was that the argumentative function of a source may change across chapters and that bibliographic correctness alone is therefore insufficient to assess the quality of source use.

The original prompt specified a five-stage procedure. First, a frame of reference was to be reconstructed for each chapter, distinguishing empirical from epistemological and methodological claims. Second, all sources actually used were to be mapped through a source cartography recording citation frequency, type of use and the specific claims supported. Third, each source was to be classified separately for each chapter according to its argumentative dependence — low, medium or high — after which all sources with medium or high dependence were to receive an individual analysis of their functional contribution and potential vulnerabilities. Fourth, a separate distributive analysis was required, covering citation frequencies, numbers of unique sources, cross-chapter centrality and the distribution of primary versus review literature, classical versus recent literature, and empirical versus philosophical sources. Finally, these analyses were to be integrated into a system-level diagnosis of the dissertation's epistemological balance.

The prompt also contained explicit validation rules. In particular, every analysed source first had to appear in the source cartography, each medium- or high-dependence classification had to correspond to an analytical block, unknown citation frequencies had to be identified as estimates, and violation of these requirements was to result in the analysis being declared invalid.

I.2 Difference between the Ex Ante Protocol and Its Actual Execution

The generated audit did **not fully execute this ex ante protocol**. Several of its central analytical principles were followed, but some procedural requirements were omitted or modified during model execution.

Most importantly, no systematic citation-frequency inventory was produced for every chapter. The requested distributive analysis was not generated as a separate cross-chapter analytical stage. In addition, sources performing closely related functions were sometimes evaluated jointly rather than in separate analytical blocks, despite the explicit requirement that every medium- or high-dependence classification should generate a corresponding individual analysis. Finally, the model did not invoke the prescribed invalidation rule after departing from these procedural requirements.

These deviations are relevant because the original prompt was deliberately designed to constrain the analysis. The resulting output demonstrates that the presence of detailed

procedural instructions in a prompt cannot itself be treated as evidence that those instructions have been executed literally. A language model may preserve the substantive logic of an analytical task while compressing, combining or omitting procedural elements. Verification of model output must therefore include verification of **procedural compliance**, rather than being confined to the plausibility or quality of the resulting text.

The discrepancy has deliberately not been concealed by retrospectively rewriting the original prompt. The ex ante prompt is retained as a documentary artefact because the difference between requested procedure and realised procedure forms part of the methodological evidence generated by this case.

I.3 Ex Post Reconstruction of the Procedure Actually Applied

Because the completed source audit remains analytically coherent despite its incomplete compliance with the original protocol, the procedure actually performed can be reconstructed ex post. This reconstruction does not claim that the original prompt was followed in full. It specifies the analytical operations that can be identified in the resulting audit itself.

The first stage consisted of a **chapter-specific reconstruction of the frame of reference**. For each chapter, its function within the dissertation was identified and the principal empirical, methodological and epistemological claims were distinguished. Source use was consequently assessed relative to the argumentative task of the individual chapter rather than relative to the dissertation as a whole.

The second stage consisted of a **classification of argumentative dependence**. Sources actually used within each chapter were classified as having low, medium or high dependence. High dependence was assigned where a source carried a central claim or substantially structured the reasoning; medium dependence where a source made a substantive but replaceable contribution; and low dependence where its function was primarily illustrative, contextual, historical or supplementary. Classification remained chapter-specific, allowing the same source to perform different functions in different parts of the dissertation.

The third stage consisted of a **targeted evaluation of sources with medium or high argumentative dependence**. The analysis examined both their positive contribution and their potential vulnerabilities. Positive contributions included empirical coverage, theoretical precision, historical positioning and methodological legitimation. Potential vulnerabilities included conceptual overloading, interpretative distortion, selective reading, loss of context, inappropriate reliance on authority and substitution of citation for argument. Sources performing closely related argumentative functions were occasionally assessed together where separate treatment would have fragmented a single analytical issue. Each assessment resulted in a proportional judgement, such as *predominantly robust*, *mixed but functional* or *robust but interpretatively strained*.

The fourth stage consisted of a **chapter-level system diagnosis**. Individual source assessments were integrated to determine whether source use was cumulatively reinforcing, whether vulnerabilities displayed a systematic direction, and whether the relationship between empirical substantiation and theoretical interpretation remained proportionate. The resulting diagnoses were qualitative rather than numerical and were explicitly restricted to source use within the relevant chapter.

I.4 Methodological Significance of the Discrepancy

The difference between the ex ante protocol and the ex post reconstruction is not merely an administrative irregularity. It exposes a methodological feature of AI-assisted knowledge production that is directly relevant to this research programme. Prompt specification and model execution are distinct stages. A detailed prompt can constrain model behaviour, but it does not guarantee literal procedural compliance.

This has consequences for the status of AI-generated analytical work. Evaluation cannot stop at the question of whether the resulting analysis appears coherent or substantively plausible. It must also examine whether the model performed the operations it was instructed to perform. In this case, the substantive core of the intended method survived: source use was assessed by chapter, argumentative dependence was differentiated, heavily weighted sources received critical scrutiny, and chapter-level diagnoses were produced. What did not survive intact was the more formal audit architecture surrounding those operations.

The ex post protocol therefore serves two purposes. It provides an accurate methodological account of the analysis actually presented in Appendix 2, while the preserved ex ante prompt documents the more elaborate procedure originally intended. The discrepancy between the two is itself informative: **a prompt records an instruction, not proof of its execution**. In AI-assisted research, procedural traceability consequently requires comparison of requested operations with realised output rather than reliance on the prompt alone.

I.5 Original Dutch text of the prompt:

Just for forensic purposes, the original prompt (in the Dutch language) – the one that had not been fully executed by the Language Model - is presented as text in the following box

Prompt – Bronnenverificatie als hoofdstuk-gebonden epistemisch systeem

Analyseer het gebruik van bronnen in onderstaande tekst als een **gedistribueerd epistemisch systeem**.

⚠ Uitgangspunt: één centrale literatuurlijst ≠ éénmalig brongebruik.
Bronnen kunnen per hoofdstuk een **andere argumentatieve functie** hebben.

Stap 0 – Referentiekader (max. 150 woorden)

Formuleer:

< wat is de centrale these van dit proefschrift(onderdeel) en hoe wordt deze per hoofdstuk opgebouwd? >

Maak expliciet:

- welke claims empirisch zijn
- welke claims epistemologisch/methodologisch zijn

Dit fungeert als anker voor alle verdere analyse.

Stap 1 – Broncartografie (verplicht, eerst uitvoeren)

Breng **feitelijk brongebruik in kaart per hoofdstuk**.

Voor elk hoofdstuk:

[Hoofdstuk X]

Per bron (alleen indien daadwerkelijk gebruikt in dit hoofdstuk):

[Volledige referentie]

- Frequentie van expliciete vermelding: (aantal of schatting)
- Type inzet: (empirisch bewijs / theoretisch kader / methodologisch / historisch / retorisch)
- Concrete claim(s) die door deze bron worden ondersteund

⚠ Regels:

- Alleen bronnen opnemen die **daadwerkelijk in de tekst voorkomen**
- Eén bron mag in meerdere hoofdstukken voorkomen (met verschillende functies)

Stap 2 – Lokale afhankelijkheidsanalyse (per hoofdstuk)

Binnen elk hoofdstuk:

Bepaal per bron de **argumentatieve afhankelijkheid** volgens:

Hoog als ≥3 criteria gelden:

1. Claim valt weg zonder bron
2. Bron structureert redenering
3. Meerdere kernpassages steunen erop
4. Geen alternatieven besproken

Middel: 1–2 criteria
Laag: illustratief/contextueel gebruik

⚠ Cruciaal:

- Classificeer **per hoofdstuk**, niet globaal
- Dezelfde bron kan dus verschillende classificaties krijgen

Stap 3 – Gerichte evaluatie (balansanalyse; alleen middel/hoge afhankelijkheid)

Per hoofdstuk, per relevante bron:
[Volledige referentie] – (Hoofdstuk X) – Afhankelijkheid: (middel/hoog)

Analyseer geïntegreerd:

(A) Functionele bijdrage (positieve component)
- Welke specifieke claim wordt door deze bron mogelijk gemaakt of versterkt?
- In welke zin verhoogt de bron de **epistemische robuustheid** van het betoog?
    - (bijv. empirische dekking, theoretische precisie, historische inbedding, methodologische legitimatie)

(B) Kwetsbaarheden (negatieve component)
Beoordeel — maar uitsluitend in relatie tot het concrete gebruik:
1. Conceptuele overbelasting
2. Interpretatieve vertekening
3. Autoriteitsmisbruik
4. Selectieve lezing
5. Contextverlies
6. Substitutie van argument door verwijzing

(C) Balansinschatting (verplicht)
Geef een expliciete weging:
- Overwegend robuust gebruik
- Gemengd (functioneel maar kwetsbaar)
- Overwegend problematisch gebruik

⚠ Deze classificatie moet volgen uit (A) + (B), niet los daarvan.
⚠ Geen bullet points in de uiteindelijke analyse; dit zijn analytische richtlijnen.

Stap 4 – Distributieve analyse (verplicht)

Voer een **kwantitatieve/structurele analyse** uit:

- Frequentieverdeling van brongebruik per hoofdstuk
- Aantal unieke bronnen per hoofdstuk
- Hergebruik van sleutelbronnen (cross-chapter centrality)
- Verhouding:
    - primaire studies vs reviews
    - klassieke vs recente literatuur

- empirisch vs filosofisch

Indien exacte telling niet mogelijk is: werk met expliciet gemarkeerde schattingen.

Stap 5 – Systeemdiagnose (balansgericht, max. 200 woorden)

Beoordeel het proefschrift als geheel in termen van **epistemische balans**:

- Waar wordt brongebruik overtuigend en cumulatief versterkend ingezet?
- Waar ontstaan systematische kwetsbaarheden (bijv. herhaald patroon van overbelasting)?
- Is er evenwicht tussen:
  - empirische onderbouwing en theoretische interpretatie
  - diversiteit van bronnen en concentratie van epistemisch gewicht

Sluit af met een expliciete totaaloordeel:

- Overwegend robuust
- Gemengd maar controleerbaar
- Structureel kwetsbaar

Verplicht outputformat

0. Referentiekader

< these + structuur >

1. Broncartografie per hoofdstuk

[Hoofdstuk X]
< lijst zoals gespecificeerd >

2. Afhankelijkheidsclassificatie per hoofdstuk

[Hoofdstuk X]
< alle bronnen met laag/middel/hoog >

3. Analyse middel/hoge afhankelijkheid

< per bron per hoofdstuk >

4. Distributieve analyse

< kwantitatieve/structurele evaluatie >

5. Systeemdiagnose

< integrale evaluatie >

Validatieregels (streng)

- Geen bron mag worden geanalyseerd zonder dat hij eerst in de cartografie voorkomt
- Classificatie moet **per hoofdstuk** gebeuren
- Aantal analyseblokken = aantal middel/hoge classificaties
- Indien bronfrequentie onbekend is → expliciet markeren als schatting
- Bij schending van deze regels: **analyse ongeldig verklaren**