

Student Supplement (SS) for Chapters 3-14: Additional Worked Examples of CERIC

In this Student Supplementary section, we provide additional worked examples illustrating how to identify the CERIC elements across five disciplinary areas. These examples are particularly useful when used alongside Chapters 3-14. We encourage learners to use these examples to develop an understanding of the varieties of CERIC elements across disciplines and how to find them in scientific articles. Instructors are encouraged to use these examples in classroom instruction and as practice assignments.

SS Chapter 3: Examples for Claims

Here, we provide worked examples for identifying Claims (see Chapter 3) in the titles and abstracts of papers across different disciplines. In each case, we indicate the Claim(s) in **boldface type** and underline the sentence builders, indicating what kind of Claim it is. We also discuss nuances as appropriate.

SS 3.1 Astronomy

Example 1: Abbott et al. (2016)

This highly cited paper reported the first detection of gravitational waves with the Laser Interferometer Gravitational-Wave Observatory (LIGO) and led to the award of the 2017 Nobel Prize in Physics to Weiss, Barish, and Thorne "for decisive contributions to the LIGO detector and the observation of gravitational waves" (The Royal Swedish Academy of Sciences, 2017). This is a classic example of a Discovery Claim in action.

Title: "**Observation of Gravitational Waves from a Binary Black Hole Merger**"

Abstract:

"On September 14, 2015, at 09:50:45 UTC the two detectors of the Laser Interferometer Gravitational-Wave Observatory simultaneously observed a transient gravitational-wave signal. The signal sweeps upward in frequency from 35 to 250 Hz, with a peak gravitational-wave strain of 1.0×10^{-21} . It matches the waveform predicted by general relativity for the inspiral and merger of a pair of black holes and the ringdown of the resulting single black hole. The signal was observed with a matched-filter signal-to-noise ratio of 24 and a false-alarm rate estimated to be less than 1 event per 203 000 years, corresponding to a significance greater than 5.1σ . The source lies at a luminosity distance of 410_{-180}^{+160} Mpc corresponding to a redshift $z = 0.09_{-0.04}^{+0.03}$. In the source frame, the initial black hole masses are $36_{-4}^{+5} M_{\odot}$ and $29_{-4}^{+4} M_{\odot}$, and the final black hole mass is $62_{-4}^{+4} M_{\odot}$, with $30_{-0.5}^{+0.5} M_{\odot} c^2$ radiated in gravitational waves. All uncertainties are defined by 90% credible intervals. **These observations demonstrate the existence of binary stellar-mass black hole systems. This**

is the first direct detection of gravitational waves and the first observation of a binary black hole merger.”

There are two main Claims in this paper:

1. **“This is the first direct detection of gravitational waves and the first observation of a binary black hole merger”**: This is a classic **Discovery Claim**. The first detection of two phenomena, gravitational waves from a binary black hole merger. In fact, one could argue that these are two separate Claims that are tightly linked, since gravitational waves can be detected from systems other than binary black holes (e.g., Abbott et al. 2017). Still, we could only detect a black hole merger through its gravitational waves, since such systems do not emit light. In any case, this combined discovery is clearly the paper’s primary Claim.
2. **“These observations demonstrate the existence of binary stellar-mass black hole systems”**: This is also a **Discovery Claim**, although of secondary importance to detecting gravitational waves. Before this paper, there was evidence of individual stellar-mass black holes (most discovered through a technique called microlensing) and stellar-mass black holes in binary systems with another normal star (known as X-ray binaries because the black hole accretes matter from the other star, and the matter emits X-rays as it falls into the black hole). However, binary black holes do not emit light and can only be detected by the ripples they produce in spacetime as they orbit each other - these are the gravitational waves. Hence, this paper is able to make this secondary Claim, which represents new knowledge based on a research question (do binary stellar-mass black hole systems exist?) based on the Evidence (gravitational wave detection) and Reasoning (consistent with binary stellar-mass black hole merger, and hence preceded by an unmerged binary system).

It is tempting to interpret the second Claim as an Implication. The existence of a binary stellar-mass black hole merger would seem to *imply* the existence of binary stellar-mass black hole systems. However, Implications are always forward-looking, such as: What can be done next? Where do we go from here? In the abstract above, the statement “demonstrate the existence” is supported by the research presented and the article's Evidence and Reasoning. It is, thus, another Claim made in the article.

Example 2: Bean et al. (2010)

This paper presents a clear example of a **Refutation Claim**, having refuted prior evidence of a giant exoplanet orbiting a nearby low-mass star.

Title: **“The Proposed Giant Planet Orbiting VB 10 Does Not Exist”**

Abstract:

“We present high-precision relative radial velocities of the very low mass star VB 10 that were obtained over a time span of 0.61 years as part of an ongoing search for planets around stars at the end of the main sequence. The radial velocities were measured from high-resolution near-infrared spectra obtained with the CRIRES instrument on the Very Large Telescope, using an ammonia gas cell. The typical internal precision of the measurements is 10 m s^{-1} . These data exhibit no significant variability and are essentially constant at a level consistent with the measurement uncertainties. Therefore, we do not detect the radial-velocity variations of VB 10 that are expected due to the presence of an orbiting giant planet, as recently proposed by Pravdo & Shaklan based on apparent astrometric perturbations. In addition, we do not confirm the $\sim 1 \text{ km s}^{-1}$ radial velocity variability of the star tentatively detected by Zapatero Osorio and colleagues using lower-precision measurements. Our measurements rule out planets with $M_p > 3 M_{\text{Jup}}$ and the orbital period and inclination suggested by Pravdo & Shaklan at better than 5σ confidence. **We conclude that the planet detection claimed by Pravdo & Shaklan is spurious on the basis of this result.** Although this work yields a non-detection, it demonstrates the potential of using ammonia cell radial velocities to detect planets around very low-mass stars.”

In this case, the Claim is clear in both the title and at the end of the abstract and is unambiguously a refutation. The authors presented new, more precise data to support this Claim (Evidence), thereby avoiding a marginal null result that would conflict with a marginal positive result (Reasoning). We also note that several declarative statements in the abstract could serve as Claims in another paper. The phrase "we do not detect the radial velocity variations of VB 10 expected due to the presence of an orbiting giant planet similar to that recently proposed by Pravdo & Shaklan" sounds like a Refutation Claim in its own right, but it is in fact part of the Reasoning that supports the much more general Claim that the planet doesn't exist. Additionally, "The typical internal precision of the measurements is 10 m s^{-1} " may be a Claim for a paper assessing the precision of a new instrument or methodology. Here, it is also part of the Reasoning that supports the Claim. Finally, the last sentence appears to be a declarative statement, “demonstrates the potential” that might be taken as a Claim, particularly as it is supported by the precision presented as Evidence. However, “potential” is the key term that indicates that this is instead an element of Implication. No Evidence of other planets orbiting low-mass stars was presented in this article; hence, this is a forward-looking statement on future work enabled by the ammonia cell technique.

SS 3.2 Biology

Example 1: Karikó et al. (2005)

This research article reported that modifications to nucleosides in RNA can suppress its function in mammalian cells and may influence how the immune system recognizes foreign cells or dead tissue. The method developed in this study was foundational to the development of mRNA vaccines, including the COVID-19 vaccines by Moderna and Pfizer-

BioNTech, the Ebola vaccine by Moderna, and several other vaccines in development targeting HIV, cytomegalovirus, rabies, RSV, influenza A, and chikungunya.

Title: “**Suppression of RNA recognition by Toll-like receptors: The impact of nucleoside modification** and the evolutionary origin of RNA”

Abstract:

“DNA and RNA stimulate the mammalian innate immune system through activation of Toll-like receptors (TLRs). DNA containing methylated CpG motifs, however, is not stimulatory. Selected nucleosides in naturally occurring RNA are also methylated or otherwise modified, but the immunomodulatory effects of these alterations remain untested. **We show that RNA signals through human TLR3, TLR7, and TLR8, but incorporation of modified nucleosides m5C, m6A, m5U, s2U, or pseudouridine ablates activity.** Dendritic cells (DCs) exposed to such modified RNA express significantly less cytokines and activation markers than those treated with unmodified RNA. DCs and TLR-expressing cells are potently activated by bacterial and mitochondrial RNA, but not by mammalian total RNA, which is abundant in modified nucleosides. **We conclude that nucleoside modifications suppress the potential of RNA to activate DCs.** The innate immune system may therefore detect RNA lacking nucleoside modification as a means of selectively responding to bacteria or necrotic tissue.”

Both the title and the two highlighted sentences in the abstract refer to the same primary Claim, that modifications to the nucleosides in RNA (i.e., the A, T, G, U "letters" along the genetic helix) suppress the potential of messenger RNA (mRNA) to activate the immune system's dendritic cells. The sentence builders "We show that...()" and "We conclude that...()" help guide us to this Claim despite a forest of field-specific jargon. This Claim is a **Correlation Claim** because it shows the relationship between two variables: substituting RNA nucleosides and significantly reduced dendritic cell activity. Because the causal mechanism of this relationship is not established, this paper does not present a Cause-Effect Claim.

The unhighlighted part of the title, "the evolutionary origin of RNA," has the makings of a seemingly important second Claim or sub-claim. However, scrutiny of the abstract indicates that this is an Implication, as the final sentence includes the phrase "may therefore detect," which is speculative and suggests a path for future work. We would need to examine the text more closely, particularly the discussion or summary sections, to confirm this.

Finally, note that while the title states the claim, it does so indirectly, without sentence-initial markers. A clearer and more direct title might be "Nucleoside modification is associated with suppression of RNA recognition by Toll-like receptors." Authors often defer clarity for a tighter or more memorable title.

Example 2: Polack et al. (2020)

This study exemplifies a Cause-and-Effect Claim that confirmed the safety and efficacy of an early mRNA COVID-19 vaccine, released in the year the pandemic began.

Title: “**Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine**”

Abstract:

“BACKGROUND: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and the resulting coronavirus disease 2019 (COVID-19) have afflicted tens of millions of people in a worldwide pandemic. Safe and effective vaccines are needed urgently.

METHODS: In an ongoing multinational, placebo-controlled, observer-blinded, pivotal efficacy trial, we randomly assigned persons 16 years of age or older in a 1:1 ratio to receive two doses, 21 days apart, of either placebo or the BNT162b2 vaccine candidate (30 µg per dose). BNT162b2 is a lipid nanoparticle–formulated, nucleoside-modified RNA vaccine that encodes a prefusion stabilized, membrane-anchored SARS-CoV-2 full-length spike protein. The primary end points were the efficacy of the vaccine against laboratory-confirmed Covid-19 and safety.

RESULTS: A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of COVID-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; **BNT162b2 was 95% effective in preventing COVID-19** (95% credible interval, 90.3 to 97.6). Similar vaccine efficacy (generally 90 to 100%) was observed across subgroups defined by age, sex, race, ethnicity, baseline body-mass index, and the presence of coexisting conditions. Among 10 cases of severe COVID-19 with onset after the first dose, 9 occurred in placebo recipients and 1 in a BNT162b2 recipient. The safety profile of BNT162b2 was characterized by short-term, mild-to-moderate pain at the injection site, fatigue, and headache. **The incidence of serious adverse events was low and was similar in the vaccine and placebo groups.**

CONCLUSIONS: **A two-dose regimen of BNT162b2 conferred 95% protection against COVID-19 in persons 16 years of age or older. Safety over a median of 2 months was similar to that of other viral vaccines.** (Funded by BioNTech and Pfizer; ClinicalTrials.gov number, NCT04368728.)”

The highlighted text identifies two primary Claims in this study about the mRNA vaccine, BNT162b2, which is found to be (1) effective at 95% protection against the COVID-19 virus and (2) safe with a similar profile to other viral vaccines. The first Claim is a **Cause-and-Effect Claim** because it arises from a randomized clinical trial involving 43,548 participants randomized to receive either a vaccine injection or a placebo. Randomized clinical trials are considered the “gold standard” of causative evidence in biomedical sciences and are required for new medication and vaccine approvals by the FDA. Unlike Correlation Claims

relating one variable to another, Cause-Effect Claims in medicine (and other fields) must meet the highest validation standards through very large sample sizes, strict randomization procedures, duration, and accurate data collection.

The second Claim is an example of an empirical (or Evidence-based) Claim about comparative safety. The phrase “similar profile to other viral vaccines” explicitly compares the vaccine’s safety outcomes to a reference class (other viral vaccines). This type of Claim:

- Relies on benchmarking against known safety profiles.
- Implies equivalence or non-inferiority rather than superiority.
- Requires data from both the new vaccine and established vaccines.

In drug regulatory frameworks, this is a very specific subtype of empirical claim, often called a **non-inferiority safety Claim**.

There are also two statements in the abstract that, on first pass, might appear to be Claims but are better categorized as Reasoning supporting the main Claims. The sentence starting "Similar vaccine efficacy (generally 90 to 100%) was observed across subgroups..." describes how different populations respond to the vaccine. This statement clearly supports the overall efficacy and is part of the Reasoning, although knowing how well the vaccine works across populations also has important implications for public health policy. In addition, the phrase "The safety profile of BNT162b2 was characterized by short-term, mild-to-moderate pain at the injection site, fatigue, and headache" describes the specific reactions people may experience after vaccination and is important for planning vaccination processes. However, this statement is ultimately a Reasoning step supporting the vaccine's side effects and overall safety.

We note that the sentence builders pointing to the Claims are vague in this abstract and, as before, absent in the title (i.e., prefacing a "There is..." at the front of the title would have clarified matters). However, we are helped by the journal's abstract structure, which separates the Background, Methods, Results, and Conclusions. Our main Claims are clearly found in both the Results and Conclusions.

SS 3.3 Chemistry

Example: Figueroa et al. (2006)

This paper details a **Method Claim** for transforming one type of chemical, acid chlorides, into another, organic nitriles, via a reaction process that at the time employed a novel use of a sodium salt. The study described several key intermediate compounds and additional regeneration steps involved in the process, thereby supporting the method’s validity.

Title: “**A Nitridoniobium(V) Reagent That Effects Acid Chloride to Organic Nitrile Conversion:** Synthesis via Heterodinuclear (Nb/Mo) Dinitrogen Cleavage, Mechanistic Insights, and Recycling”

Abstract:

“**The transformation of acid chlorides (RC(O)Cl) to organic nitriles (RC:N) by the terminal niobium nitride anion [N:Nb(N[Np]Ar)₃]- ([1a-N]-, where Np = neopentyl and Ar = 3,5-Me₂C₆H₃) via isovalent N for O(Cl) metathetical exchange is presented.** Nitrido anion [1a-N]- is obtained in a heterodinuclear N₂ scission reaction employing the molybdenum trisamide system, Mo(N[R]Ar)₃ (R = t-Bu, 2a; R = Np, 2b), as a reaction partner. Reductive scission of the heterodinuclear bridging N₂ complexes, (Ar[R]N)₃Mo-(μ-N₂)Nb(N[Np]Ar)₃ (R = t-Bu, 3b; R = Np, 3c) with sodium amalgam provides 1 equiv each of the salt Na[1a-N] and neutral N:Mo(N[R]Ar)₃ (R = t-Bu, 2a-N; R = Np, 2b-N). Separation of 2-N from Na[1a-N] is readily achieved. Treatment of salt Na[1a-N] with acid chloride substrates in tetrahydrofuran (THF) furnishes the corresponding organic nitriles concomitant with the formation of NaCl and the oxo niobium complex O:Nb(N[Np]Ar)₃ (1a-O). Utilization of ¹⁵N-labeled ¹⁵N₂ gas in this chemistry affords a series of ¹⁵N-labeled organic nitriles establishing the utility of anion [1a-N]- as a reagent for the ¹⁵N-labeling of organic molecules. Synthetic and computational studies on model niobium systems provide evidence for the intermediacy of both a linear acylimido and niobacyclobutene species along the pathway to organic nitrile formation. High-yield recycling of oxo 1a-O to a niobium triflate complex appropriate for heterodinuclear N₂ scission has been developed. Specifically, addition of triflic anhydride (Tf₂O, where Tf = SO₂CF₃) to an Et₂O solution of 1a-O provides the bistriflate complex, Nb(OTf)₂(N[Np]Ar)₃ (1a-(OTf)₂), in near quantitative yield. One-electron reduction of 1a-(OTf)₂ with either cobaltocene (Cp₂Co) or Mg(THF)₃(anthracene) provided the monotriflate complex, Nb(OTf)(N[Np]Ar)₃ (1a-(OTf)), which efficiently regenerates complexes 3b and 3c when treated with the molybdenum dinitrogen anions [N₂Mo(N[t-Bu]Ar)₃]- ([2a-N₂]-) or [N₂Mo(N[Np]Ar)₃]- ([2b-N₂]-), respectively.”

The highlighted text shows that the primary **Method Claim** concerns transforming one chemical type into another via a multi-step process that employs a novel use of a sodium salt. The Method Claim is right at the front of the abstract, and its nature is implied by the simple sentence builder "is presented." The remainder of the abstract details the various steps of the Method in a language that only a chemist could love. Yet, even without knowing what the chemical terms or compounds are, we can recognize that this paper is conveying a new way of doing something (the research gap) and provides both Evidence (the various chemical compounds and reaction steps) and Reasoning (the outcomes of these reactions) that support the novelty and efficacy of the new Method. An experienced chemist will pick up on the importance of the anionic nitridoniobium(V) reagent being employed as its sodium salt, the modeling and quantitative verification of the intermediates involved in this transformation process (e.g., linear acylimido and niobacyclobutene species), and the efficient recycling of some chemical complexes (e.g., 3b and 3c), all of which provide detailed

Reasoning to demonstrate that their Method works and is verifiable. While most of this paper is very dense in jargon, a non-expert reader can still identify its main Claim.

SS 3.4 Earth Science

Example 1: Enfield, Mestas-Nuñez, & Trimble (2001)

Title: **“The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US”**

Abstract:

“North Atlantic sea surface temperatures for 1856–1999 contain a 65–80 year cycle with a 0.4 °C range, referred to as the Atlantic Multidecadal Oscillation (AMO) by Kerr [2000]. AMO warm phases occurred during 1860–1880 and 1940–1960, and cool phases during 1905–1925 and 1970–1990. **The signal is global in scope, with a positively correlated co-oscillation in parts of the North Pacific, but it is most intense in the North Atlantic and covers the entire basin there. During AMO warmings most of the United States sees less than normal rainfall, including Midwest droughts in the 1930s and 1950s.** Between AMO warm and cool phases, Mississippi River outflow varies by 10% while the inflow to Lake Okeechobee, Florida, varies by 40%. The geographical pattern of variability is influenced mainly by changes in summer rainfall. The winter patterns of interannual rainfall variability associated with El Niño-Southern Oscillation are also significantly changed between AMO phases.”

The primary Claim is a **Relational Claim** (also known as a Pattern or Correlational Claim), describing the relationship between the Atlantic Multidecadal Oscillation (AMO) phase and patterns in U.S. rainfall and river flows, particularly during warm AMO phases.

“The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US”

As we read further into the abstract, it tells us what the relationship is:

“The signal is global in scope, with a positively correlated co-oscillation in parts of the North Pacific, but it is most intense in the North Atlantic and covers the entire basin there. During AMO warmings most of the United States sees less than normal rainfall, including Midwest droughts in the 1930s and 1950s.”

In other words, the key relationship links two sets of variables—AMO phase and U.S. rainfall/river flow patterns—by emphasizing co-variation over time and differences between phases, rather than by centering on a cause-and-effect mechanism. The title phrase “its relation to” and the abstract language “positively correlated,” “pattern of variability,” and “variability associated with” cue that relational structure.

Example 2: Shean et al. (2020)

Title: **“A systematic, regional assessment of high mountain Asia glacier mass balance”**

Abstract:

“High-mountain Asia (HMA) constitutes the largest glacierized region outside of the Earth's polar regions. Although available observations are limited, long-term records indicate sustained HMA glacier mass loss since ~1850, with accelerated loss in recent decades. Recent satellite data capture the spatial variability of this mass loss, but spatial resolution is coarse and some estimates for regional and HMA-wide mass loss disagree. To address these issues, we generated 5,797 high-resolution digital elevation models (DEMs) from available sub-meter commercial stereo imagery (DigitalGlobe WorldView-1/2/3 and GeoEye-1) acquired over HMA glaciers from 2007 to 2018 (primarily 2013–2017). We also reprocessed 28,278 ASTER DEMs over HMA from 2000 to 2018. We combined these observations to generate robust elevation change trend maps and geodetic mass balance estimates for 99% of HMA glaciers between 2000 and 2018. We estimate total HMA glacier mass change of -19.0 ± 2.5 Gt yr⁻¹ (-0.19 ± 0.03 m w.e. yr⁻¹). We document the spatial pattern of HMA glacier mass change with unprecedented detail, and present aggregated estimates for HMA glacierized sub-regions and hydrologic basins. Our results offer improved estimates for the HMA contribution to global sea level rise in recent decades with total cumulative sea-level rise contribution of ~0.7 mm from exorheic basins between 2000 and 2018. **We estimate that the range of excess glacier meltwater runoff due to negative glacier mass balance in each basin constitutes ~12–53% of the total basin-specific glacier meltwater runoff.** These results can be used for calibration and validation of glacier mass balance models, satellite gravimetry observations, and hydrologic models needed for present and future water resource management.”

This abstract provides a lot of Context and a series of Evidence statements. If we keep scanning, buried near the bottom, we arrive at what appears to be a classic **Modelling Claim**.

“We estimate that the range of excess glacier meltwater runoff due to negative glacier mass balance in each basin constitutes ~12–53% of the total basin-specific glacier meltwater runoff. “

This information completes the idea presented in the authors' stated title and represents a logical conclusion based on the evidence and analysis presented. We could double-check our thinking by stitching these pieces together to see if they make sense:

[We modelled] **“a systematic, regional assessment of high mountain Asia glacier mass balance [and based on that model] we estimate that the range of excess glacier meltwater runoff due to negative glacier mass balance in each basin constitutes ~12–53% of the total basin-specific glacier meltwater runoff. “**

Indeed, the Modelling Claim is now much clearer.

SS 3.5 Physics

Example: Novoselov et al. (2004)

This highly cited paper describes the electrical properties of a single-atom layer of carbon, graphene, synthesized by the authors, who demonstrated that it is a stable nanomaterial with high tensile strength, electrical conductivity, and transparency. This work led to a revolution in solid-state physics and the development of various applications of graphene in electronics, batteries, and other materials. It also earned the authors Konstantin Novoselov and Andre Geim the Nobel Prize in Physics in 2010 for "groundbreaking experiments regarding the two-dimensional material graphene."¹

Title: "Electric Field Effect in Atomically Thin Carbon Films"

Abstract:

"We describe monocrystalline graphitic films, which are a few atoms thick but are nonetheless stable under ambient conditions, metallic, and of remarkably high quality. The films are found to be a two-dimensional semimetal with a tiny overlap between valence and conductance bands, and they exhibit a strong ambipolar electric field effect such that electrons and holes in concentrations up to 10^{13} per square centimeter and with room-temperature mobilities of $\sim 10,000$ square centimeters per volt-second can be induced by applying gate voltage."

The primary Claim in this paper is a **Phenomenon Claim** that describes the physical and electrical properties of a new material, graphitic films. Classifying this as a Discovery Claim is tempting, given that this is the first paper to describe graphene. However, the paper focuses on material properties, as indicated by the sentence starter "We describe..." and by the subsequent list of material properties. The phenomenon focus is indicated in the title, which focuses on the material's electrical properties. While the latter part of this abstract dips into pretty significant Physics jargon, we nevertheless get a clear sense of the Claim from the plain-language declarative statement at the start of the abstract.

¹ <https://www.nobelprize.org/prizes/physics/2010/press-release/>

SS Chapter 4: Additional Examples for Evidence

Here, we provide worked examples of identifying Evidence (see Chapter 4) in the Abstracts and Method sections of papers across different disciplines. In each case, we indicate the Evidence in **boldface type** and underline the sentence builder that points to it. We also discuss nuances as appropriate. These examples are useful for classroom discussions and practice.

SS 4.1 Astronomy

Example: Clowe et al. (2006)

This highly-cited paper has as its title an eye-catching Claim: "A Direct Empirical Proof of the Existence of Dark Matter." Dark matter is a hypothesized form of invisible matter that exerts a gravitational force but does not interact with light. Remarkably, there appears to be 5-6 times more dark matter in the universe than the normal matter we interact with² (i.e., atoms, molecules, planets, stars, ice cream), but we still don't know what dark matter is. This paper reported observations revealing the existence of dark matter, spatially distinct from ordinary matter, in a merging galaxy cluster known as 1E0657-558, also called the "bullet cluster." As the claim emphasizes "empirical proof", the Evidence is critical in this study, and is featured prominently in the abstract:

"We present **new weak-lensing observations of 1E 0657-558 (z=0.296)**, a unique cluster merger, that enable a direct detection of dark matter, independent of assumptions regarding the nature of the gravitational force law. Due to the collision of two clusters, the dissipationless stellar component and the fluid-like X-ray-emitting plasma are spatially segregated. By **using both wide-field ground-based images and HST/ACS images of the cluster cores**, we create gravitational lensing maps showing that the gravitational potential does not trace the plasma distribution, the dominant baryonic mass component, but rather approximately traces the distribution of galaxies. An 8 σ significance spatial offset of the center of the total mass from the center of the baryonic mass peaks cannot be explained with an alteration of the gravitational force law and thus proves that the majority of the matter in the system is unseen."

The primary Evidence of this paper is presented twice: first, by introducing "**new weak-lensing observations**" at the start, which is effectively a map of the light distortions on distant galaxies caused by the gravitational bending of spacetime by dark matter. This opening sentence indicates that the Evidence is primarily **observational**, as is common in astronomy. The authors further specify in the abstract the specific type of observations obtained, "**ground-based images and HST/ACS images.**"

² <https://home.cern/science/physics/dark-matter>

The Evidence is further detailed in Table 1 ("Optical Imaging Sets"), which specifies details about when and where the images were obtained and their quality; and at the end of Section 2 "Methodology and Analysis": "...four sets of optical images shown in Table 1 and the VLT image set used in Clowe et al. (2004)." Finally, the authors display their optical imaging data alongside another important piece of evidence: an X-ray image obtained with the Chandra X-ray Observatory, which traces the "dominant baryonic component" or "normal mass" in this system. The weak-lensing map they produce is overlaid on both images, and it is ultimately the spatial offset between the distortions (tracing dark matter) and the X-ray emission (tracing normal matter) that enables the authors to substantiate their claim. The repeated Emphasis on observations throughout this paper demonstrates its importance in supporting the study's Claim.

SS 4.2 Biology

Example: Long et al. (2020)

In the early days of the COVID-19 pandemic, medical practitioners didn't know why some people exposed to the COVID-19 virus experienced significant and even life-threatening symptoms, including severe respiratory illness, fever, and cough, while others showed mild or no symptoms whatsoever. Understanding the variance in response to virus exposure was an important step in both planning treatment for those infected as well as predicting the spread of virus among so-called "silent spreaders". This study found important differences between asymptomatic and symptomatic patients, including that the former shed virus longer and had a weaker immune response than the latter.

The study focused on a group of people who were exposed to the SARS-CoV-2 virus but did not exhibit any clinical symptoms related to infection; from the abstract:

"We studied 37 asymptomatic individuals in the Wanzhou District who were diagnosed with RT-PCR-confirmed SARS-CoV-2 infections but without any relevant clinical symptoms in the preceding 14 d and during hospitalization."

The main text then goes on to detail some of the tests undertaken to track the infection, immune response, and symptoms of this group:

"A complete blood count, blood biochemistry, coagulation function, liver and renal function and infection biomarkers were measured upon admission (Supplementary Table 2) to monitor the potential disease progression."

In addition, the individuals had chest computed tomography to image the lung scarring that had come to be associated with COVID infection, nasal swabbing to measure the viral load over time, measurement of immunoglobulins (IgG and IgM) in the blood, and a series of

cytokines and chemokines that track inflammation in the body. Some of these measurements were made during both the "acute" stage (shortly after infection) and "convalescent" phase (8 weeks after hospital discharge) to track the development of the immune response. Since these measurements were all taken in a lab setting, this paper illustrates an example of **Experimental Evidence**.

As an example of a Correlation Claim in the medical sciences, the design of the sample is particularly important. Evidence was collected for both the asymptomatic group and for an equal number of COVID-infected individuals with relevant symptoms to determine how the populations differed. Importantly, the authors note that the symptomatic group consisted of "sex-, age-frequency- and comorbidity-matched" patients, reducing potential biases along these demographic characteristics. The authors also investigated a similarly matched "control" group of individuals who were not infected with COVID to help identify potential false positives in the comparison among the COVID-infected groups. In addition, all patients were drawn from the same Wanzhou District and studied at the same time, further reducing biases from differences in health care access or virus mutations. As such, this study's Evidence includes both experimental measurements and sample characteristics.

Finally, many of the details of the sample and the measurements obtained are provided in a supplementary "Methods" section following the main text. This additional section provides more information on the patient demographics, how data were specifically collected, the individual cytokines measured, etc. A supplement is a common feature of some articles where the main text focuses on the other CERIC elements (in particular, Claim, Implications, and Context) while leaving details in Evidence and sometimes Reasoning as a separate section for interested readers.

SS 4.3 Chemistry

Example: Yan et al. (2024)

One of the top chemistry results in 2024 reported a new approach to hydrogen extraction from ammonia. This work specifically addresses new energy sources, noting that molecular hydrogen (H_2) is "the cleanest energy source and an important industrial raw material, [and] is one of the foundations of the future society." The Methodological Claim indicated in the title, "Efficient and Rapid Hydrogen Extraction from Ammonia–Water via Laser Under Ambient Conditions without Catalyst", requires a combination of **Process and Procedure Evidence** (what was the methodology?) and **Experimental Evidence** (how did they prove their methodology was valid?). Both forms of evidence are briefly touched on in the abstract:

“Here, we demonstrate a simple, green, ultrafast, and highly efficient method for hydrogen extraction from ammonia–**water by laser bubbling in liquids (LBL) at room temperature and ambient pressure without catalyst. A maximum apparent yield of 33.7 mmol/h and a real yield of 93.6 mol/h**

were realized in a small operating space, which were far higher than the yields of most hydrogen evolution reactions from ammonia–water under ambient conditions.”

More detail is provided in the Results and Discussion section, where a basic schematic of their experimental setup is displayed as an illustration (Figure 1), and there is brief descriptions of the experimental setup ("250 mL of ammonia–water was added to a homemade reactor with an airtight quartz cylinder"), operation of the laser ("After passing through the lens, the laser is focused below the ammonia level through an optical window at the top of the reactor"), extraction of H₂ from the resulting bubbles ("carried out of the reactor by Ar flow and then washed with dilute sulfuric acid and dilute sodium hydroxide solution"), and measurement of the resulting abundance ("The hydrogen concentration is detected with gas chromatography"). This section also reports the associated temperatures in the bubbles and the hydrogen yield for different variations in experimental parameters (e.g., Air flow rate, laser energy, laser frequency, and the time over which the system is operated).

All of these details are presented quite tersely, leaving the reader perhaps unsure of key points such as how the temperatures were measured, how the chemical pathways were determined, and the type of laser used and how the pulses were timed. As in the Long et al. (2020) article, these details are presented in a Supporting Information section separate from the main body. In this case, the supplement data is quite extensive; while the main article is about 6 pages long in print (excluding references), the Supporting Information is more than twice as long and contains highly specific information on the experimental setup (including where the ammonium water and laser were sourced), assumptions used to compute the chemical pathways, and specifics on how the bubble temperatures were measured, all with associated diagrams. A movie of the laser pulsation and bubble formation is also provided. Delving into supplemental information or materials may be necessary to clarify the Evidence that supports an article's Claim.

SS 4.4. Earth Science

Example: Enfield, Mestas-Nuñez, & Trimble (2001)

Let's revisit this paper that was discussed as an example of a Relational Claim in SS 3.4. Earth's climate naturally varies across multiple timescales, from the familiar annual variations we describe as seasons to longer-term variations driven by interactions among the land, ocean, and atmosphere, or by shifts in Earth's orbit that can take decades to millennia to manifest. Enfield et al. examine one of these oscillations, the Atlantic Multidecadal Oscillation (AMO), and focus on its impact on rainfall in the United States. AMO is a very long-period oscillation with a period of about 65-70 years. As such, long-term datasets are required to see the impacts of this oscillation on regional weather conditions, indicated by the timescales listed in the abstract:

“North Atlantic sea surface temperatures for 1856–1999 contain a 65–80 year cycle with a 0.4 °C range, referred to as the Atlantic Multidecadal Oscillation (AMO) by Kerr [2000]. AMO warm phases occurred during 1860–1880 and 1940–1960, and cool phases during 1905–1925 and 1970–1990.”

This first sentence describes one of the datasets used in this study: a set of global sea surface temperature (SST) measurements spanning 1856-1999. More specifically, the authors use temperature "anomalies," which are differences between the measured and reference temperatures at a local region. The main sets of data used are described further in the article in the "Data and Methods" section:

“Our study is based on three data sets: an updated (1856-1999) version of the Kaplan et al. [1998] monthly reanalysis of **global SST anomalies (SSTA)**, **monthly rainfall over the continental United States** summarized by climate divisions (1895-1999) [National Climatic Data Center], and (as independent hydrological checks) the **records of Mississippi River outflow and the indirectly estimated inflow into Florida's Lake Okeechobee.**”

The authors further describe analyses applied to these data in order to focus on the long-period oscillation, including "applying a ten-year running mean", "linearly detrending", and defining an index based on temperature anomalies in the north Atlantic where they are largest. All of these analysis steps are designed to prepare measurements for examining correlations with rainfall and river flow data and should therefore be considered part of the Evidence, rather than Reasoning, as these steps do not directly lead to the study's Claim.

Later in the "Discussion" section of the article, the authors introduce a new set of data, height measurements corresponding to an air pressure of 500 hPa ("geopotential") that are calculated from another reference:

“To probe the explanation for the patterns we see, we calculated the composite average distributions of 500 hPa geopotential height from the NCEP/NCAR reanalysis [Kalnay et al., 1996] for two periods, 1949-1969 and 1970-1994, and subtracted the average for 1949-1999 (not shown).”

The purpose of this dataset is to evaluate changes in air pressure associated with storm formation and rainfall, and, as such, to support the claim that the AMO correlates with wet and dry phases in the continental US. Again, both the original data (from Kalnay et al., 1996) and the calculation of geopotential height should be considered further Evidence, whereas the connection between geopotential height and rainfall is part of the Reasoning.

The Evidence presented in this study is almost exclusively experimental measurements drawn from archival data, with additional analysis applied to properly identify variations on

the decadal timescale of the Atlantic Multidecadal Oscillation. As such, this paper illustrates an example of **Archival Evidence**. This study emphasizes the importance of unique datasets - a nearly 150-year record of sea surface temperature measurements, and 100-year records of rainfall and river flow - to identify and characterize patterns in nature.

SS 4.4 Physics

Example: Landy et al. (2008)

Metamaterials are synthetic materials or compounds that are engineered to have physical or optical properties not typically found in natural materials. Some famous examples of metamaterials include those with negative indices of refraction³ and those that can act like "cloaking devices" at specific wavelengths of light.⁴ In this study, the authors report the development of a metamaterial that is a highly efficient absorber ("Perfect Metamaterial Absorber"), and is an example of a claim based on the development of a new resource, tool, or instrument. To justify their claim, the authors present three forms of Evidence: the design and construction of their metamaterial (**Process and Procedure Evidence**), simulation of the metamaterial's properties (**Simulation Evidence**), and experimental measurements used to validate these properties (**Experimental Evidence**). All three elements are present in the article's abstract:

"We present the design for an absorbing metamaterial element with near unity absorbance. Our structure consists of two metamaterial resonators that couple separately to electric and magnetic fields so as to absorb all incident radiation within a single unit cell layer. **We fabricate, characterize, and analyze** a metamaterial absorber with a slightly lower predicted absorbance of 96%. This achieves a simulated full width at half maximum (FWHM) absorbance of 4% thus making this material ideal for imaging purposes. Unlike conventional absorbers, **our metamaterial consists solely of metallic elements**. The underlying substrate can therefore be chosen independently of the substrate's absorptive qualities and optimized for other parameters of interest. We detail the design and simulation process that led to our metamaterial, and our experiments demonstrate a peak absorbance greater than 88% at 11.5 GHz."

Much of the abstract is devoted to the design of the metamaterial - its structure and composition - and the abstract explicitly calls out the design, fabrication, simulation, characterization, and analysis elements of the paper to support their primary Claim. In addition, four of the five figures in the paper focus on Evidence elements: Figure 1 displays a schematic of the metamaterial, Figure 2 displays their simulation results, and Figures 3 and 4

³ https://en.wikipedia.org/wiki/Negative-index_metamaterial

⁴ https://en.wikipedia.org/wiki/Metamaterial_cloaking

compare simulations to experimental measurements. The authors also provide precise specifications for the construction of their metamaterial in the text, and illustrate their experimental design to measure absorption, reflectance, and transmission within Figure 3. As with other papers highlighting new resources, tools, or instruments, this paper should clearly focus on Evidence so that other groups can replicate and validate the properties of this new metamaterial.

SS Chapter 5: Additional Examples for Reasoning

Here, we repeat some of the examples reviewed for Reasoning in Chapter 5, and separately draw out the elements of Reasoning that develop the main logical argument, motivate the study design and Evidence, and support the Claim.

SS 5.1 Astronomy

Example: Clowe et al. (2006)

As discussed in Chapter 4 and SS 4.1, Clowe et al. (2006) claim direct empirical Evidence of the existence of dark matter by showing that there is an unseen matter spatially separated from luminous regular matter in a merging galaxy cluster called 1E0657-558, also known as the "bullet cluster."

Reasoning that develops the main logical argument: The logical framework for the main argument in this article is **Deductive Reasoning**. The authors are explicitly aiming to distinguish two theories for why galaxies don't behave as predicted by Newton's Universal Law of Gravitation: that either an unseen form of matter exists that is distinct from regular matter, or that Newton's Law is somehow modified for matter on a galactic scale.

In the Introduction, the authors establish the predictions of the two main theoretical hypotheses as the framework for their main argument:

"**[If]** In the absence of dark matter, **[then]** the gravitational potential will trace the ... X-ray plasma. **If**, on the other hand, the mass is indeed dominated by collisionless dark matter, **[then]** the potential will ... be spatially coincident with the collisionless galaxies. **Thus**, by deriving a map of the gravitational potential, one can discriminate between these possibilities."

This section provides a typical pair of *if...then* couplets for our hypotheses, followed by a test that will enable us to distinguish one from the other. The authors then go on to build up the logical chain of this argument. In the Analysis section, they detail the finding of two peaks in their mass map:

"Two major peaks are clearly visible in the reconstruction, one centered ... north of the subcluster's brightest cluster galaxy (BCG) and detected at **8 σ significance**...and one centered south ... detected at **12 σ** ."

and confirm that these peaks are spatially separated from the X-ray plasma:

"Both peaks are also offset at **~8 σ** from the center of mass of their respective plasma clouds."

Note that both of these sentences detail the significance of the measurements, in terms of both detectability and the spatial offset between mass and X-ray light, using multiples of "sigma" or uncertainty. Because the uncertainties are important in establishing significance, the authors provide supporting Reasoning of how their statistical approach to determine uncertainties in the positions (centroids):

"We estimate centroid uncertainties by repeating bootstrap samplings of the background galaxy catalog, performing a κ reconstruction with the resampled catalogs, and measuring the centroid of each peak."

The authors also address potential biases in the construction of their mass map:

"the change in κ due to the mass-sheet degeneracy should be less than 1% and any systematic effects on the centroid and skewness of the peaks are much smaller than the measured error bars."

By experimentally establishing the separation of mass and X-ray light predicted by the dark matter model for a merging pair of galaxies, Clowe et al. have deductively reasoned the existence of dark matter. Moreover, by noting that other systems show similar separations (see below), the authors also make an inductive argument that dark matter is the underlying explanation for the gravitational properties of all galaxies at the end of the Discussion section:

"The observed displacement between the bulk of the baryons and the gravitational potential proves the presence of dark matter for the most general assumptions regarding the behavior of gravity."

Reasoning that supports the study design and Evidence: In their Introduction section, Clowe et al. (2006) note define the experimental conditions by which the two competing models for dark matter can be distinguished:

"The actual existence of dark matter can **only be confirmed** either **by** a laboratory detection or, in an astronomical context, by the discovery of a system in which the observed baryons and the inferred dark matter are spatially segregated. An ongoing galaxy cluster merger is such a system."

This line of Reasoning justifies the study's observational target, the Bullet Cluster, because it is a system of merging galaxies. In addition, the authors using Reasoning to motivate the measurements needed and the expected outcome that would support their eventual Claim:

"In the absence of dark matter, the gravitational potential will trace the ... X-ray plasma. If, on the other hand, the mass is indeed dominated by collisionless dark

matter, the potential will ... be spatially coincident with the collisionless galaxies. Thus, by deriving a map of the gravitational potential, one can discriminate between these possibilities."

The authors also have numerous examples of Reasoning used to justify specific choices in their study design, which provide support for their Evidence. For example, in the Analysis section the authors describe how they convert the measured optical light of the galaxies ("luminosities") to stellar mass using a known conversion factor:

"The luminosities were converted into mass assuming ... $M/L = 2$."

The authors subsequently acknowledge the intrinsic uncertainty in this assumption, but also argue that the stellar mass is less important to the mass contained in the X-ray plasma:

"The assumed mass-to-light ratio is highly uncertain (can vary between 0.5 and 3) and depends on the history of recent star formation of the galaxies in the apertures; however even in the case of an extreme deviation, the X-ray plasma is still the dominant baryonic component in all of the apertures."

Reasoning that supports the Claim: The authors provide additional Reasoning in their Discussion section that further supports their Claim. For example, they acknowledge that "a structure unassociated with the cluster could exist and be projected onto the lensing map" but deduce that the probability of this alignment is $\approx 10^{-7}$, and hence exceedingly low. They also consider whether larger-scale structures known as cosmological filaments, the intersections of which may be the sites of galaxy cluster formation, could contaminate the mass signal. They argue that the probability of such alignment is $\ll 10^{-8}$, even lower.

The authors also bring in supporting studies; for example:

"...we mention that two other merging clusters, MS1054-03 (Jee et al. 2005) and A520 (in preparation), exhibit similar offsets between the peaks of the lensing and baryonic mass, although based on lensing reconstructions with lower spatial resolution and less clear-cut cluster geometry."

While these other studies of clusters with similar offsets between mass and light peaks may themselves be insufficient to establish the existence of dark matter (note the qualifiers and limitations in the last sentence), combined with the Clowe et al. measurements these studies paint a consistent picture of distinct dark matter and regular matter being present in the interacting galaxy systems.

Finally, the authors also consider counterfactuals; namely, that other theories (in this case modified gravity) cannot reproduce the observations:

"Any non-standard gravitational force that scales with baryonic mass will fail to reproduce these observations. The lensing peaks require unseen matter concentrations that are more massive than and offset from the plasma."

On the other hand, there is room for additional hypotheses, including a mixture of dark matter and non-standard gravity models:

"While the existence of dark matter removes the primary motivation for alternative gravity models, it does not preclude non-standard gravity."

This statement is consistent with the underlying uncertainties and limitations in the Evidence and Claim.

Critiquing the article's Reasoning: Clowe et al. (2006) present a compelling deductive argument for the existence of dark matter, but even robust studies merit scrutiny. First, their deductive reasoning depends heavily on a binary hypothesis framework: either dark matter exists or gravity behaves differently on cosmic scales. While this clean separation simplifies testing, it may obscure other hybrid possibilities, such as dark matter and modified gravity models. The authors acknowledge this, noting, "it does not preclude non-standard gravity" (p. 5), but this caveat ends there.

Second, while they quantify uncertainties in their gravitational lensing measurements—for instance, using bootstrap resampling and sigma significance values—they assume that systematic errors from lens modeling and background galaxy shape distortions are negligible. The authors state that "any systematic effects on the centroid and skewness of the peaks are much smaller than the measured error bars" (p. 3), but this conclusion is based on internal estimates. Future work with more precise instruments or different lens reconstruction algorithms might shift the inferred mass centroids or peak significance levels.

Third, the study leans on strong spatial correlations to argue that the gravitational potential aligns with the galaxies rather than the plasma. But while correlation supports the claim, it does not confirm causality. For instance, the separation between mass and plasma is inferred to result from dark matter's collisionless nature during the merger. Yet this interpretation depends on the assumption that the collision geometry is well-understood. If the merger axis is significantly inclined relative to the line of sight, projected 2D maps might misrepresent the actual 3D mass-plasma configuration.

Finally, the inductive extension of the dark matter model to other merging clusters—like MS1054–03 and A520—is suggestive but not definitive. As the authors admit, those systems "are based on lensing reconstructions with lower spatial resolution and less clear-cut cluster

geometry” (p. 4). Thus, the generalizability of the findings, while reasonable, remains speculative.

Despite these areas for improvement, Clowe et al. (2006) demonstrate a high standard of Reasoning to support an important Claim.

SS 5.2 Biology

Example: Long et al. (2020)

As discussed in Chapter 4, Long et al. (2020) investigate why some individuals exposed to SARS-CoV-2 develop severe symptoms—such as respiratory distress and fever—while others remain asymptomatic or mildly affected. This section focuses on the Reasoning that connects their immunological evidence to the Claim that asymptomatic individuals mount a weaker immune response. Unlike Clowe et al. (2006), who used Deductive Reasoning to test competing models in the previous astronomy example, Long et al. (2020) employ **Inductive Reasoning**, building general Claims from patterns in observed data.

Reasoning that supports the study design and Evidence: The authors begin by justifying the need for their study, citing Bai et al. (2020) and Hu et al. (2020) to emphasize the role of asymptomatic carriers in viral spread. This sets up the research gap: the immune profiles of these carriers remain poorly understood. Their Claim that asymptomatic individuals exhibit weaker immune responses directly addresses this gap.

To ensure meaningful comparisons in the inductive argument, they stratified participants into three well-defined groups:

1. **Symptomatic SARS-CoV-2-positive patients** confirmed via RT-PCR, serving as a positive control;
2. **SARS-CoV-2-negative individuals** confirmed via RT-PCR, serving as a negative control;
3. **Asymptomatic SARS-CoV-2-positive patients** confirmed via RT-PCR, serving as the experimental group.

All participants were **matched by sex, age distribution, and comorbidities** within the Wanzhou District region simultaneously. This careful matching is a key component of methodological reasoning: it helped minimize potential confounding variables, enhanced comparisons across groups, and strengthened the study’s internal validity.

The authors then implemented a range of standard medical diagnostic tools—RT-PCR tests, antibody assays, chest radiographs, and complete blood counts—to assess immune response. They analyzed these data using statistical methods, such as box plots and Mann-Whitney tests. Reasoning emerges in how the authors interpret statistical significance ($p <$

.05) as support for meaningful differences between groups, strengthening the inferential link from observation to conclusion.

Reasoning that supports the Claim: Long et al. (2020) draw on several lines of Reasoning to support their Claim in the discussion. **Analogical Reasoning** compares SARS-CoV-2 to biologically similar viruses like SARS-CoV and MERS-CoV, for example:

"In SARS-CoV, viral RNA was detectable...() for as long as 4 weeks after disease onset. In MERS-CoV infections, viral shedding in respiratory secretions persisted for at least 3 weeks" (p. 1203-04).

They reason other CoV viruses show prolonged RNA shedding, and thus, it is plausible that SARS-CoV-2 does as well. This analogical Reasoning helps contextualize their findings in prior MERS-CoV research that asymptomatic individuals shed the virus for a median of 19 days—longer than symptomatic individuals.

Next, **Experimental and Causal Reasoning** underpin the interpretation that a longer duration of shedding in asymptomatic individuals suggests delayed immune clearance, for example:

"IgG levels and neutralizing antibodies in a high proportion of individuals who recovered from SARS-CoV-2 infection start to decrease within 2–3 months after infection" (p. 1204)

This time frame for waning IgG immunity (i.e., humoral or antibody-based immunity) is significantly shorter than for other CoV viruses, which range from 1-3 years after infection. However, the authors qualify this interpretation. They note the presence of viral RNA does not equate to infectiousness, and RT-PCR cannot detect whether virus particles are viable. These limitations are part of a Reasoning framework that acknowledges uncertainty.

Long et al. (2020) also use **Methodological Reasoning** to explain differences across studies. They note that definitions of shedding duration, severity classification, and testing frequency all influence findings. Further **Epistemic Reasoning** helps readers understand why studies vary and why generalizations require caution.

Finally, Long et al. (2020) use **Policy Reasoning** to generate Implications. If immunity wanes quickly (i.e., 2-3 months), as their IgG data suggest, policies such as immunity passports may be premature. This Reasoning connects individual-level immunological data to societal-level decisions.

Critiquing the article's Reasoning: While Long et al. (2020) present a compelling inductive argument, the paper would benefit from additional Reasoning types. A clearer mechanistic explanation could reinforce causal Claims, for instance, by exploring *why* asymptomatic

individuals might clear the virus more slowly—perhaps due to lower levels of initial immune activation. More explicit **Probabilistic Reasoning**—such as quantifying the likelihood of reinfection—would sharpen their treatment of uncertainty. Epistemic Reasoning could be extended by comparing study design trade-offs, such as the benefits and limits of cross-sectional versus longitudinal approaches.

Still, Long et al. (2020) offer a strong example of Inductive Reasoning in biological research. Their argument builds from carefully collected observations to a cautious but consequential Claim about the immune response. Like Clowe et al. (2006), they acknowledge uncertainty and limitations and invite future studies to test the durability of their conclusions—a hallmark of strong scientific Reasoning.

SS 5.3. Chemistry

Example: Yan et al. (2024)

As discussed in Chapter 4, Yan et al. (2024) investigated a new method for hydrogen extraction from ammonia–water, called Laser Bubbling in Liquids (LBL). In this section, we analyze the **Abductive Reasoning** used in their study. This type of Reasoning begins with an unexpected observation (i.e., unusually high hydrogen yields under mild conditions) and infers the most plausible explanation (i.e., localized, transient high temperatures produced by pulsed lasers). This strategy exemplifies what is known as *inference to the best explanation*, rather than strict deduction or induction.

Reasoning that supports the study design and Evidence: Yan et al. (2024) begin by establishing the broader significance of hydrogen production and the challenge it presents:

“Hydrogen (H₂) is emphasized as a clean energy source and an important industrial raw material”, but “it is difficult to transport due to flammability and explosiveness” (p. 4864).

As a solution, they focus on *in situ* hydrogen generation using ammonia–water, which contains “17.6 wt% hydrogen” (p. 4864) but presents technical hurdles for extraction.

They highlight limitations of conventional methods:

“Thermal catalytic decomposition of ammonia is endothermic and requires high temperatures and significant energy input” (p. 4864).

Alternatives like electrocatalysis and photocatalysis “demand expensive precious metals or show low efficiency and instability” (p. 4864). To address this, the authors propose LBL as a novel, catalyst-free method.

They argue that “short, intense laser pulses ... generate cavitation bubbles with transiently high temperatures” exceeding 10,000 K (p. 4865). These bubbles enable hydrogen extraction without heating the whole system. This study design leverages “superfast heating and cooling” ($>10^{10}$ K/s), creating far-from-equilibrium conditions that prevent reverse reactions (p. 4865). Because LBL “does not require catalytic surfaces,” it avoids cost and stability issues associated with precious metals (p. 4865).

Additional empirical justification comes from their yield data:

“The apparent hydrogen yield rate was 33.7 mmol/h,” but accounting for laser “on” time (0.36 ms per hour), “the real hydrogen yield rate reached 93.6 mol/h” (p. 4867).

This stark contrast supports the hypothesis that brief, extreme conditions drive the reaction.

Reasoning that supports the Claim: *Mechanistic Reasoning* dominates their discussion. The authors state:

“When a laser is applied to ammonia–water, the molecules at the focus point are excited and ionized ...() Notably, these small bubbles have extreme internal transient temperatures ...() convenient microenvironments for hydrogen extraction” (p. 4865).

Indeed, they cite prior studies on cavitation (Refs. 10–13, 15–20) that back this Claim. While actual bubble temperatures are inferred, not directly measured, the reasoning is supported by well-established physics.

Comparative reasoning explains why LBL outperforms traditional methods. They explain:

“Compared to traditional catalytic reactions ...() the LBL technique is more similar to the ‘batch-by-batch’ method” (p. 4865).

They add another comparison that discrete laser pulses avoid steady-state back-reactions common in catalytic systems.

Thermodynamic and kinetic reasoning appears in the claim that “rapid cooling moves the reaction away from equilibrium ...() preventing the reverse reaction” (p. 4865–66). Once hydrogen forms, it escapes before equilibrium restores, improving yield. DFT calculations strengthen the mechanistic claims. The authors report:

“ $\text{H}\cdot/\text{OH}\cdot/\text{NH}_2\cdot$ radicals are considered the active species in the LBL process” (p. 4868).

These species form above 5,000 K, supporting the claim that radicals—not just heat—drive hydrogen production.

Feasibility Reasoning suggests future industrial potential. “LBL is a simple, clean, and efficient method ...() solar light can be efficiently converted into a laser” (p. 4869). The authors cite >60% conversion efficiency and kilowatt-scale solar-pumped lasers.

Critiquing the article’s Reasoning: Yan et al. (2024) offer a strong abductive case, but several areas could be strengthened. First, direct evidence of bubble temperatures and intermediate species remains elusive. While Mechanistic Reasoning is compelling, it rests on indirect observations. A future study using ultrafast spectroscopy or imaging would strengthen this link.

Second, **Probabilistic Reasoning** is limited. While they discuss yield and timing, they do not quantify variability across runs or estimate uncertainties in radical concentrations or reaction intermediates. A clearer statistical treatment would improve their precision.

Third, the economic Feasibility Reasoning is promising but vague. They note that “energy consumption is an issue” and suggest solar pumping as a solution (p. 4869), but omit specific cost comparisons or life-cycle analyses.

Still, the paper illustrates how Abductive Reasoning works in chemistry: observations lead to explanatory hypotheses, which are justified through comparative, mechanistic, and feasibility arguments. Like Clowe et al. (2006) and Long et al. (2020), the authors logically connect the Evidence to their Claims, address limitations, and propose directions for future study. The result is a layered and persuasive argument that demonstrates *inference to the best explanation* in experimental chemistry.

SS 5.4 Earth Science

Example: Enfield, Mestas-Nuñez, & Trimble (2001)

As discussed in Chapter 4, Enfield et al. (2001) investigated how the Atlantic Multidecadal Oscillation (AMO), a 65–70-year ocean temperature cycle, influences rainfall and hydrology across the United States. In this section, we examine how the authors use a combination of **Abductive and Inductive Reasoning**.

They began with the unexplained oscillation identified by Schlesinger and Ramankutty (1994) and inferred the best explanation for its influence on climate: a modulating relationship with regional hydrological variability. Through pattern-seeking across multiple datasets, the authors supported their Claim using data-driven generalizations.

Reasoning that supports the study design and Evidence: The authors begin by identifying a limitation in existing models:

“Schlesinger and Ramankutty (1994) identified a 65–70 year North Atlantic surface temperature oscillation ...() but existing models fail to replicate [it]” (p. 2076).

This establishes their abductive hypothesis—that the AMO, although not fully modeled, may still play a key role in precipitation variability. They clearly articulate the research gap:

“the direct impact on precipitation (especially in the continental U.S.) and its relationship with interannual factors ...() are not well understood” (p. 2076).

To address this, Enfield et al. (2021) propose that “natural multidecadal variability could explain observed trends in U.S. precipitation” (p. 2076). This formulation, grounded in unexplained observations, exemplifies **Abductive Reasoning**.

However, the study’s structure follows an inductive path. The authors gather empirical Evidence to test whether AMO phases align with patterns in rainfall and river flows. The authors used long-term datasets—Kaplan et al. (1998) SST anomalies (1856–1999), U.S. rainfall (1895–1999), and hydrologic records (e.g., Mississippi River outflow, Lake Okeechobee inflow). They applied a 10-year running mean to linearly detrended data:

“We focus on the multidecadal character ...() by applying a ten-year running mean to linearly detrended time series of all the data” (p. 2077).

This Reasoning isolates low-frequency variability associated with the AMO, reducing ENSO-style interannual noise. **Comparative Reasoning** strengthens the study’s internal validity, for example:

“records of Mississippi River outflow and ...() Lake Okeechobee inflow ...() are compared to the appropriate area-weighted rainfall accumulations” (p. 2077).

They also reason that river discharge patterns should match rainfall anomalies to indicate data integrity. Indeed, the authors argue that “flow strongly mimics rainfall signals” (p. 2077), supporting the data’s reliability, noting the limitation that evapotranspiration data are missing,

Reasoning that supports the Claim: The AMO index itself is defined inductively:

“We index the AMO with a ten-year running mean of Atlantic SSTA north of the equator ...() The temporal variations reproduce the phases and periodicity previously ascribed to the AMO” (p. 2078).

High spatial correlations with North Atlantic SST anomalies justify its use. The central Claim—that AMO phases affect regional U.S. rainfall—is supported by spatial correlation analyses, for example:

“The correlations of ...() rainfall with the AMO index display a robust continental-scale pattern dominated by negative correlations ...() with positive correlations in the Pacific Northwest, the northeast, and Florida” (p. 2079).

The Mississippi basin shows reduced rainfall during AMO warm phases while Florida shows the opposite. These patterns suggest a modulating role for the AMO. Correlations between ENSO and rainfall also shift with AMO phase, for example:

“During the 1930–1960 warm phase of the AMO the rainfall had a significant negative correlation with NINO-3.4 ...() during the cool phases ...() correlations were insignificant” (p. 2079).

This line of Reasoning implies that the AMO modulates ENSO’s teleconnections, creating nonstationary climate relationships. **Model-based Reasoning** supports a broader Implication, for instance:

“global temperatures can be expected to be greater than they would be based only on greenhouse [gases] ...() attempts to anticipate ...() may prove inaccurate if the models do not reproduce the AMO variability” (p. 2080).

Thus, failing to account for the AMO could mislead climate projections. Finally, **Policy Reasoning** appears in water management contexts:

“net average annual inflow to Lake Okeechobee was about double ...() [in the positive AMO phase] ...() translates into a near complete reversal in water management priorities” (p. 2080).

This implies that long-term water policies must account for decadal variability.

Critiquing the article’s Reasoning: Enfield et al. (2001) present a compelling, multi-layered argument, and several Reasoning elements could be strengthened. First, while correlation patterns are strong, they are not causative. The authors acknowledge this, noting the “mid-1940s anomaly” and suggesting that “other oceanic influences” may be at play (p. 2079). Second, while they apply Monte Carlo phase randomization (citing Ebisuzaki, 1997) to handle serial correlation, the absence of mechanistic modeling limits explanatory power. They do not simulate how AMO SST anomalies generate rainfall shifts—this is **Observational Reasoning**, not dynamic modeling. Third, **Probabilistic Reasoning** is largely absent. The authors rely on significance thresholds but do not quantify uncertainties in the AMO index or rainfall correlation metrics across basins.

Nonetheless, the article exemplifies a productive fusion of Abductive and Inductive Reasoning. Like the previous disciplinary examples, the authors build a careful sequence:

they identify an observational pattern, hypothesize a plausible explanation, gather empirical support, compare across variables, and acknowledge caveats. The result is a layered and persuasive argument about the AMO's role in modulating continental-scale rainfall.

SS 5.5 Physics

Example: Landy et al. (2008)

As discussed in Chapter 4, Landy et al. (2008) investigated a metallic metamaterial that achieves near-unity absorption of microwave radiation. They use **Deductive Reasoning**, starting with theoretical principles—namely, that perfect absorption requires impedance matching and minimal transmission. They designed a structure predicted to satisfy these conditions and tested it experimentally, progressing from theoretical prediction to empirical validation.

Reasoning that supports the study design and Evidence: The study begins by identifying a key gap:

“perfect absorption at subwavelength thickness had not yet been experimentally demonstrated” (p. 1).

Prior approaches struggled to align electric and magnetic resonances or required thicker materials. The authors elaborate on the reasons for their choice:

“designer electromagnetic materials are an ideal platform for the investigation of novel emergent physical phenomena while also holding great promise for future applications” (p. 1).

Landy et al. (2008) apply electromagnetic theory—specifically, Maxwell's equations and impedance matching conditions—to justify their study design. Their approach centers on constructing a metamaterial unit cell with a split-ring resonator and a cut-wire structure. They reason:

“electric resonance arises from the cut-wire geometry” and “magnetic resonance arises from current loops in the split-ring” (p. 2).

When these are tuned to resonate at the same frequency, the metamaterial achieves an effective impedance (Z) matched to free space (Z_0), minimizing reflection. They also anticipate and address potential limitations, for example:

“uncertainties in geometrical tolerances, fabrication imperfections, or material losses could degrade perfect absorption” (p. 2).

Nonetheless, they argue that precise geometric tuning minimizes these effects. This Reasoning is grounded in boundary conditions and effective medium theory, establishing a clear, logical chain from theory to design.

Reasoning that supports the Claim: The experimental findings support the theoretical prediction. The authors report:

“measured transmission (S₂₁) is nearly zero at resonance ...() reflection (S₁₁) is also nearly zero” (p. 3).

They apply to classical microwave theory, which indicates that absorption $A = 1 - R - T$. With both R and T near zero, “ $A \approx 0.99$ ” (p. 3). This demonstrates near-perfect absorption. Further, they reason:

“the surface current distributions confirm the expected electric and magnetic resonances coincide in frequency” (p. 3), validating their design logic.

They also present simulation data that closely matches their measurements, reinforcing confidence in the physical interpretation. The authors conclude that:

“perfect absorption is the logical outcome of impedance matching ...() and co-located resonances at the same frequency” (p. 4).

Since their structure is only $\lambda/35$ thick, traditional quarter-wave resonance cannot explain this achievement because it depends entirely on engineered electromagnetic response.

They acknowledge measurement uncertainty and bandwidth limitations, for example:

“the absorption is narrowband and subject to material losses” (p. 4).

Nonetheless, they argue that the design is generalizable to other frequencies with appropriate tuning.

Critiquing the article’s Reasoning: Landy et al. (2008) offer a strong deductive argument, but several areas could be strengthened. First, while the absorption is near-unity, it is also narrowband. The authors acknowledge, “This design is not broadband ...() [though] the approach is generalizable” (p. 4). Second, the reliance on idealized geometry could be sensitive to fabrication variability. Slight errors in the gap or thickness of components could shift resonant frequency or reduce Q-factor. This issue is acknowledged but not deeply explored. Third, the reasoning assumes ideal free-space matching, but the real measurement setup includes substrate effects and finite boundaries. The authors do not provide a full

uncertainty analysis or explore performance under oblique incidence or polarization variations.

Still, their logic is robust. They begin with a clear theoretical prediction, derive a testable design, and validate the results with empirical data. Like the other disciplinary examples, Landy et al. (2008) used multiple forms of Reasoning—including electromagnetic theory, measurement, and simulation—to construct a tightly reasoned argument. The result is a demonstration of how Deductive Reasoning can guide discovery in experimental physics.

SS Chapter 6: Additional Examples for Implications

We revisit the examples presented for Implications in Chapter 6, illustrating how to identify this element in the Abstract and Discussion/Conclusions sections. We also highlight cases where Implications are not explicitly discussed, and how to infer them from the subsequent literature.

SS 6.1 Astronomy

Example: Penzias and Wilson (1965)

In their landmark paper, “A Measurement of Excess Antenna Temperature at 4080 Mc/s,” Arno Penzias and Robert Wilson discovered Cosmic Microwave Background (CMB) radiation while working with a radio antenna at Bell Labs. They detected a faint, uniform signal, which they described as an excess in antenna temperature not attributable to known sources. Later, this finding was interpreted as the afterglow of the Big Bang.

Authors’ stated Implications: Penzias and Wilson do not state any Implications of their research in the original article! You may indeed find that some articles, particularly those that present small-scale observational or experimental results, or define a new process or tool, may lack Implications. Nevertheless, given that Penzias and Wilson shared the 1978 Nobel Prize in Physics, there must be broader Implications to the work.

Broader Implications: By exploring some of the work that cites the Penzias and Wilson (1965) study, including a summary from the Nobel Prize in Physics,⁵ we can deduce the paper’s broader Implications and summarize them as:

1. The detection of excess antenna temperature revealed the existence of CMB radiation, and these measurements provided strong evidence supporting the Big Bang theory, an example of a **Theoretical Implication**.
2. The uniformity of the CMB across the sky supported the idea of an early hot and dense state of the universe, which has since expanded and cooled, further extending the Theoretical Implication.
3. This finding significantly altered our understanding of cosmology and the evolution of the universe, and example of a **Paradigm Shift Implication**.

In brief, Penzias and Wilson's (1965) discovery of excess antenna temperature directly led to major advances in physical cosmology, which emerged as a rigorous, experimental scientific discipline. This work transformed cosmology from a largely theoretical endeavor to a rich interplay between experiment and theory, giving this article important **New Research Field Implications**. Measurement of the CMB has since evolved to probe the spatial structure

⁵ The Nobel Prize Foundation, retrieved 22 February 2026:
<https://www.nobelprize.org/prizes/physics/1978/summary/>

needed to form the first galaxies; spatial correlations that reveal the relative abundances of normal matter, dark matter, and dark energy in the Universe; and, most recently, polarization studies to search for evidence of the universe's early inflation; all of these represent **Future Research Implications**. This rich set of Implications, while not discussed in the original paper, explains why Penzias and Wilson's study has had such a major impact on cosmology.

SS 6.2 Biology

Example: Watson and Crick (1953a,b)

In their landmark pair of papers entitled "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid" and "Genetical Implications of the Structure of Deoxyribonucleic Acid," James Watson and Francis Crick discovered the structure of DNA, which was revealed as the double helix model, where DNA consists of two strands twisted around each other. This breakthrough, built on the fundamental X-ray crystallography work of Rosalyn Franklin and Raymond Gosling, helped explain how genetic information is stored, replicated, and passed on from generation to generation. DNA structure underpins the field of molecular biology, which describes the flow of genetic information in living systems.

Authors' stated Implications: Watson and Crick (1953a) published a brief description of the molecular structure of DNA and then published the complete model (1953b). The first report gives two Implications:

1. **"This structure has novel features which are of considerable biological interest"** (p. 737, paragraph 1). This is a well-framed, and perhaps highly understated, **Future Research Implication**, since the description of DNA's structure would indeed enable many new directions in biological research.
2. **"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material"** (p. 737, paragraph 12). Again, in circumspect style, Watson and Crick propose the Holy Grail of biology research at the time: how genetic material copies itself. This statement can be interpreted as a **Theoretical Implication** since it proposes a broader idea – replication – based on the measured structure.

In the second paper, Watson and Crick (1953b) present the complete model, and they rain down many important Implications:

3. **"Genetical Implications of the Structure of Deoxyribonucleic Acid"** (p. 964, title). Watson and Crick's proposed structure of DNA had so many implications that they knew about at the time (and certainly many more would follow), they included the word "Implications" in the title.
4. **"Despite these uncertainties, we feel that our proposed structure for deoxyribonucleic acid may help to solve one of the fundamental biological**

problems—the molecular basis of the template needed for genetic replication” (p. 966, paragraph 8). Here, they are more explicit in stating that their model provides a framework for understanding DNA replication as a molecular process, a clear **Theoretical Implication** since the mechanism of replication is not worked out in this paper.

5. **“Even if it is correct, it is clear from what we have said that much remains to be discovered before the picture of genetic duplication can be described in detail. What are the polynucleotide precursors? What makes the pair of chains unwind and separate? What is the precise role of the protein? Is the chromosome one long pair of deoxyribonucleic acid chains, or does it consist of patches of the acid joined together by protein?”**(p. 966, paragraph 7). As they conclude the article, Watson and Crick identify essential, unanswered questions as **Future Research Implications**. We now know that researchers at the time considered these Implications carefully and transformed them into research questions that, in turn, led to new significant discoveries.

Broader Implications: Over the years, scientists have built on Watson and Crick’s (1953) work to elucidate the mechanisms of DNA replication, transcription, and translation, thereby laying the groundwork for new fields, technologies, and tools. Their discovery of the structure of DNA laid the foundation for the modern framework of the genetic code, gene function, and mutation. This understanding revolutionized biology and gave birth to the field of molecular biology. Major impacts of this work included the development of genetic engineering and gene therapy, the Human Genome Project, and advances in biotechnology. Their work also laid the foundation for the emergence of the fields of genetics and genomics, making **New Research Fields** part of the paper’s further Implications.

SS 6.3 Chemistry

Example: Brenner and Lerner (1992)

Sydney Brenner and Richard Lerner’s landmark paper, “Encoded Combinatorial Chemistry,” introduced a technique that revolutionized drug discovery by quickly generating and screening vast libraries of chemical compounds. This paper introduced the concept of DNA-encoded combinatorial chemistry for drug discovery, and DNA-encoded combinatorial libraries have since become a core tool in medicinal chemistry.

Authors’ stated Implications: Brenner and Lerner (1992) anticipate many changes their method will generate, and devote the final two paragraphs of their paper to these Implications:

1. ...() **Our method permits the study of the efficacy of combinatorial associations of diverse chemical units without the necessity of either synthesizing them one at a time or knowing their interactions in advance. It also allows easy identification of**

the most effective molecules through a common nucleic acid sequencing method. Once the chemical polymers are decoded, more precise questions about critical interactions and conformations can be asked by reversion to classical chemical methods. Further, we expect that many receptors will interact with sets of related but not identical chemical entities such that major clues as to critical interactions can be deduced from the shared features of the sets” (p. 5383, paragraph 4). Brenner and Lerner detail a series of **Practical Implications** in how their method can be used to explore new chemical combinations, as well **Future Research Implications** in exploring decoded chemical polymers are decoded and related chemical entities.

2. **“Our method also provides a method of amplification, again by exploiting a common procedure of nucleic acid hybridization... By coupling genetics and the versatility of organic chemical synthesis we have extended the range of analysis to chemicals that are not themselves part of biological systems”** (p. 5383, paragraph 5). This text illustrates both **Methodological and Strengthen Current Practice Implications**, as Brenner and Lerner describe how their technique can efficiently amplify the presence of non-biological chemical species using existing DNA synthesis techniques, in a way not possible at this time.

Initially, Brenner and Lerner (1992) focused on the immediate benefits of their DNA-encoded combinatorial chemistry, emphasizing its ability to streamline the identification of effective molecules via nucleic acid sequencing and amplification. They predicted that their method would enable rapid screening and improve precision in understanding molecular interactions. Over time, these capabilities revolutionized drug discovery.

Broader Implications: This work expanded the possibilities for high-throughput screening and drug discovery by introducing a method for generating vast libraries of chemical compounds using DNA as a molecular barcode. This enabled rapid identification of bioactive compounds from large libraries, thereby revolutionizing the discovery of pharmaceutical leads. The work’s major impacts include significantly accelerating drug discovery and developing more efficient methods for screening compounds for specific biological activities. The method has also led to the discovery of numerous new drugs and therapeutic compounds. This work also laid the foundation for the emergence of the fields of DNA-encoded libraries and chemical biology, which have grown into vital areas in drug development and molecular medicine. The principles described in this research article also paved the way for advancements in personalized medicine and the rational design of therapeutic agents based on specific genetic markers, which tips their stated Methodological and Practical Implications into **New Research Field and Paradigm Shift Implications** in drug development.

SS 6.4 Earth Science

Example: Manabe & Wetherald (1967)

The landmark paper by Syukuro Manabe and Richard. T. Wetherald, "The Earth's Radiation Budget and Climate Change," developed the first climate models that simulated the effect of increased carbon dioxide on Earth's temperature, linking atmospheric CO₂ concentrations to global warming. This groundbreaking research was the first sound climate model that enabled accurate predictions of climate change, including the first computational assessment of the magnitude of atmospheric emissions and water vapor feedback. The model provided the first scientific basis for understanding anthropogenic climate change, and its results are still valid today.

Authors' stated Implications: Manabe and Wetherald (1967) explicitly state the Implications of their climate model as estimates of the effects of CO₂ and atmospheric moisture on global temperatures in several places:

1. **“According to our estimate, a doubling of the CO₂ content in the atmosphere has the effect of raising the temperature of the atmosphere (whose relative humidity is fixed) by about 2 °C.”** (p. 241, abstract paragraph 4). Manabe and Wetherald's model makes specific predictions of temperature increase if CO₂ abundance goes from 300 parts per million (ppm) measured at that time to 600 ppm, a good example of an **Extending Trends Implication**. Given that at the time of this writing CO₂ abundance has exceeded 420 ppm,⁶ and global temperatures have risen by over 1 °C,⁷ their prediction is highly salient to our climate future.
2. **“4) Doubling the existing CO₂ content of the atmosphere has the effect of increasing the surface temperature by about 2.3 °C for the atmosphere with the realistic distribution of relative humidity and by about 1.3 °C for that with the realistic distribution of absolute humidity.”** (p. 254, paragraph 5). Manabe and Wetherald also make predictions about the intersection of humidity and temperature increase, demonstrating how the former can modify the potential range of future global temperature rise.
3. **“5) A five-fold increase of stratospheric water vapor from the present value of 3 x 10⁻⁶ gm gm⁻¹ of air causes an increase of surface equilibrium temperature of about 2.0 °C, when the vertical distribution of relative humidity is fixed. Its effect on the equilibrium temperature of the stratosphere is larger than that of the troposphere.”** (p. 254, paragraph 6). Again, focusing on the intersection of humidity and temperature, Manabe and Wetherald's model produces a higher rise in temperature in the stratosphere (upper atmosphere) compared to the troposphere

⁶ Scripps Institute of Oceanography CO₂ Program, retrieved 22 February 2026: <https://scrippsco2.ucsd.edu/>

⁷ NOAA Climate.gov, retrieved 22 February 2026: <https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature>

(near the surface of the Earth), a prediction made possible by the model's simulation of the entire atmosphere.

Manabe and Wetherald's (1967) model initially provided critical estimates of how increased CO₂ and atmospheric moisture would affect global temperatures, projecting potential temperature rises attributable to these factors. While these Extending Trends Implications focused on the immediate outcomes of CO₂ increases and changes in humidity, the predictions also formed the basis for much broader Implications of how climate models can be used in the future.

Broader Implications: Manabe and Wetherald's (1967) model laid the foundation for the field of climate modeling by projecting future climate change and quantifying the value of climate sensitivity. The work also had significant **Policy Implications**, as climate models today inform global climate policy discussions, climate impact assessments, and efforts to mitigate climate change through international agreements such as the 2015 Paris Agreement. Their work also enabled the emergence of the fields of climate science and global climate modeling, giving the study **New Field of Research Implications**.

SS 6.5 Physics

Example: Bell (1964)

The landmark paper by John S. Bell (1964), "Bell's theorem and the foundations of quantum mechanics," presented a theorem showing that no physical theory based on local hidden variables could reproduce all the predictions of quantum mechanics, thereby fundamentally challenging classical notions of reality. This theorem challenges the concept of locality or local realism, in which a particle is only influenced by factors within a causal distance limited by the speed of light. Bell's paper directly addresses early work by Albert Einstein and his colleagues that critiqued quantum theory based on the Einstein-Podolsky-Rosen (EPR) paradox, in which quantum particles can be "entangled" and influence each other beyond the causal distance, which Einstein described as "spooky action at a distance."

Authors' stated Implications: Bell (1964) articulates two main Implications of the work:

1. **"That particular interpretation has indeed a grossly non-local structure. This is characteristic, according to the result to be proved here, of any such theory which reproduces exactly the quantum mechanical predictions"** (p. 195, Introduction paragraph). Bell acknowledges here that any theory fully replicating quantum mechanical predictions, must inherently involve non-locality, a **Theoretical Implication** for interpreting quantum mechanics.
2. **"Of course, the situation is different if the quantum mechanical predictions are of limited validity. Conceivably, they might apply only to experiments in which the settings of the instruments are made sufficiently in advance to allow them to**

reach some mutual rapport by exchange of signals with velocity less than or equal to that of light. In that connection, experiments of the type proposed by Bohm and Aharonov [6], in which the settings are changed during the flight of the particles, are crucial.” (p. 199, Conclusion). He also acknowledges that alternative explanations may arise if quantum predictions are valid only under specific conditions, making experiments such as those proposed by Bohm and Aharonov essential for testing these limits. This instance is a good example of a **Replication or Methodological Implication**, as it suggests new experiments to validate the theorem, experiments subsequently conducted in the 1980s that validated Bell’s theorem and demonstrated non-locality in quantum interactions.

Broader implications: Initially, a theoretical challenge to local realism evolved into a cornerstone of quantum mechanics. Over time, the Implications expanded beyond philosophical debates, influencing practical advancements in quantum cryptography, quantum computing, and secure communication technologies, examples of **Practical Implications**. Concurrently, this work laid the foundation for the new field of quantum information science or quantum computing, an example of a New Research Field Implication.

SS Chapter 7: Additional Examples for Context

Here, we provide worked examples of identifying Context (see Chapter 7) in the titles and abstracts of papers across different disciplines. Additional detail should be sought in the introductory sections of these articles, but we can often get the broad outline of the Introductory Funnel (see Section 7.2.4) in the initial article elements. In each case, we present the Context(s) in **boldface** and underline the Claim, signaling where the Introductory Funnel begins and ends. We also identify specific aspects of the Introductory Funnel model in these examples.

SS 7.1 Astronomy

Example: Nandakumar et al. (2023)

This research article explored the impact of telecommunications satellites on astronomical observing, arguing that the telecommunications satellite *BlueWalker 3* sometimes outshines nearly all stars (99%) visible from Earth. This satellite is part of a growing trend toward such bright satellite constellations which have the potential to contaminate or complicate astronomical observations. Thus, the research team suggests that future launches should undergo impact assessments. Nonetheless, this abstract stands out for its lengthy introduction establishing an Introductory Funnel, and it includes citations that function as a mini-literature review.

Title: “The high optical brightness of the BlueWalker 3 satellite”

Abstract:

“**Large constellations of bright artificial satellites in low Earth orbit pose significant challenges to ground-based astronomy (ref. 1). Current orbiting constellation satellites have brightnesses between apparent magnitudes 4 and 6, whereas in the near-infrared Ks band, they can reach magnitude 2 (ref. 2). Satellite operators, astronomers and other users of the night sky are working on brightness mitigation strategies (refs. 3,4). Radio emissions induce further potential risk to ground-based radio telescopes that also need to be evaluated.** Here we report the outcome of an international optical observation campaign of a prototype constellation satellite, AST SpaceMobile’s BlueWalker 3. BlueWalker 3 features a 64.3 m² phased-array antenna as well as a launch vehicle adaptor (LVA; ref 5). The peak brightness of the satellite reached an apparent magnitude of 0.4. This made the new satellite one of the brightest objects in the night sky. Additionally, the LVA reached an apparent V-band magnitude of 5.5, four times brighter than the current International Astronomical Union recommendation of magnitude 7 (refs. 3,6); it jettisoned on 10 November 2022 (Universal Time), and its orbital ephemeris was not publicly released until 4 days later. The expected build-out of constellations with hundreds of thousands of new bright objects (ref. 1) will make active satellite tracking and avoidance strategies a necessity for ground-based telescopes.”

Let's break down the start of this abstract using the Introductory Funnel model:

- **“Large constellations of bright artificial satellites in low Earth orbit pose significant challenges to ground-based astronomy (ref. 1).”**

The first sentence establishes the broad Context (ground-based astronomy) and key theme (bright artificial satellites in low Earth orbit) of the article.

- **“Current orbiting constellation satellites have brightnesses between apparent magnitudes 4 and 6, whereas in the near-infrared Ks band, they can reach magnitude 2 (ref. 2).”**

The second sentence lays out the scale of the problem, specifying brightnesses of the satellites at visible and near-infrared bands.

- **“Satellite operators, astronomers and other users of the night sky are working on brightness mitigation strategies (refs. 3,4).”**

The third sentence summarizes the work to date on the problem to develop brightness mitigation strategies. Since the authors do not identify specific shortcomings of this work in the abstract, we would need to hunt further into the main text to identify the “research gap” in the literature that provides the rationale for the study.

- **“Radio emissions induce further potential risk to ground-based radio telescopes that also need to be evaluated.”**

This statement notes a second related and potentially significant problem, the radio emissions of these satellites and their risk to ground-based radio telescopes.

- **“Here we report the outcome of an international optical observation campaign of a prototype constellation satellite, AST SpaceMobile’s BlueWalker 3.”**

Finally, Introductory Funnel concludes with the Claim, a **Phenomenon Claim** that describes observations of the brightness of a new satellite, BlueWalker 3.

Looking backward, we can see that the Claim of the paper provides quantitative support that bolsters prior measurements of satellite measurement, and the supports the need to improve and implement mitigation strategies.

SS 7.2 Biology

Example: Barber et al. (2024)

This study explored geographical variations in the sexual selection of birds of various species, and the impacts of climate and climate change on bird breeding behaviors. Like Nandakumar et al. (2023), this article’s abstract establishes an Introductory Funnel that helps situate its Claim in the broader research.

Title: “Climate and ecology predict latitudinal trends in sexual selection inferred from avian mating systems”

Abstract:

“Sexual selection, one of the central pillars of evolutionary theory, has powerful effects on organismal morphology, behaviour, and population dynamics. However, current knowledge about geographical variation in this evolutionary mechanism and its underlying drivers remains highly incomplete, in part because standardised data on the strength of sexual selection is sparse even for well-studied organisms. Here, we use information on mating systems—including the incidence of polygamy and extra-pair paternity—to estimate the intensity of sexual selection in 10,671 (>99.9%) bird species distributed worldwide. We show that avian sexual selection varies latitudinally, peaking at higher latitudes, although the gradient is reversed in the world’s most sexually selected birds—specialist frugivores—which are strongly associated with tropical forests. Phylogenetic models further reveal that the strength of sexual selection is explained by temperature seasonality coupled with a suite of climate-associated factors, including migration, diet, and territoriality. Overall, these analyses suggest that climatic conditions leading to short, intense breeding seasons, or highly abundant and patchy food resources, increase the potential for polygamy in birds, driving latitudinal gradients in sexual selection. Our findings help to resolve longstanding debates about spatial variation in evolutionary mechanisms linked to reproductive biology and also provide a comprehensive species-level data set for further studies of selection and phenotypic evolution in the context of global climatic change.”

Let’s break down this introduction using the Introductory Funnel model:

- **“Sexual selection, one of the central pillars of evolutionary theory, has powerful effects on organismal morphology, behaviour, and population dynamics.”**

The first sentence establishes the broad context (evolutionary theory), the key theme (sexual selection), and the scope of impacts (organismal morphology, behavior, and population dynamics) of the problem this article focuses on.

- **“However, current knowledge about geographical variation in this evolutionary mechanism and its underlying drivers remains highly incomplete, in part because standardised data on the strength of sexual selection is sparse even for well-studied organisms.”**

A few things are happening in this sentence. It tells us about the shortcomings of current knowledge, notably on geographical variation in sexual selection and what may drive such variations; and explains the origin of the shortcoming in the sparsity of standardized data.

- “Here, we use information on mating systems—including the incidence of polygamy and extra-pair paternity—to estimate the intensity of sexual selection in 10,671 (>99.9%) bird species distributed worldwide.”

This at first appears to be a Claim; however, if we scan to the next sentence, we see that the authors used this information to generate another, more salient Claim, indicating that this is describing the Evidence (incidence of polygamy and extra-pair paternity) and Reasoning (>99.9% of bird species worldwide) that supports the main Claim in the following sentence.

- “We show that avian sexual selection varies latitudinally, peaking at higher latitudes, although the gradient is reversed in the world’s most sexually selected birds—specialist frugivores—which are strongly associated with tropical forests.”

This statement is the more impactful prediction Claim because it applies the authors' data on mating systems to the geography of sexual selection. The title also states this information, further supporting the idea that it is the paper’s main Claim.

The Claim directly addresses the Introductory Funnel presented in the abstract by resolving the known shortcoming in our understanding of geographical variation in sexual selection among birds through a large and highly complete sample, revealing both broad and species-dependent trends.

SS 7.3 Chemistry

Example: Becke (1993)

This seminal article in physical chemistry was part of a series of papers that explored improvements to density-functional thermochemistry (DFT), a method for computing the thermochemical properties (e.g., heats of formation and reaction energies) of molecules and solid-state materials by statistically modeling their electron densities. This work has been crucial for quantum chemistry calculations of large molecular systems, with impacts in material science, geology, and other fields.

The article provides a terse Context in the abstract to set up the main Claim.

Title: “Density-functional thermochemistry. III. The role of exact exchange”

Abstract:

“Despite the remarkable thermochemical accuracy of Kohn–Sham density-functional theories with gradient corrections for exchange–correlation [see, for example, A. D. Becke, *J. Chem. Phys.* 96, 2155 (1992)], we believe that further improvements are unlikely unless exact-exchange information is considered. Arguments to support this view are presented, and a semiempirical exchange–correlation functional containing local-spin-density, gradient, and exact-exchange terms is tested on 56 atomization energies, 42 ionization potentials, 8 proton affinities, and 10 total atomic energies of first- and second-row systems. This functional performs significantly better than previous functionals with gradient corrections only, and fits experimental atomization energies with an impressively small average absolute deviation of 2.4 kcal/mol.”

Let's break down the start of this abstract using the Introductory Funnel model:

- **“Despite the remarkable thermochemical accuracy of Kohn–Sham density-functional theories with gradient corrections for exchange-correlation [see, for example, A. D. Becke, *J. Chem. Phys.* 96, 2155 (1992)], we believe that further improvements are unlikely unless exact-exchange information is considered.”**

The first sentence quickly establishes the broad context (thermochemistry) and key theme (Kohn-Sham density functional theory), and establishes the main research gap, that this theory won't improve “unless exact-exchange information is considered.” The rationale for why improvement is needed is left vague here because they are part of the main findings of the article (see next sentence). This single sentence efficiently conveys the Context and motivation for the article.

- Arguments to support this view are presented...

The paper's first Claim appears next, indicating that it explains theoretically why the existing DFT model could not be improved without the inclusion of exact exchange information.

- This functional performs significantly better than previous functionals with gradient corrections only ...

The paper's second and primary Claim demonstrates that inclusion of exact exchange correlation functionals significantly improves the agreement between theoretical calculations and measurements of the atomization energies and ionization potentials, and proton affinities of several chemical species.

These two Claims address the stated research gap in our ability to accurately compute accurate thermochemical quantities, arguing both why the existing DFT model was unable to improve, and demonstrating how a DFT that includes exact-exchange information can produce more accurate values. This brief article, cited over 100,000 times, laid the groundwork for quantum chemistry calculations thereafter.

SS 7.4 Earth Science

Example: Shean et al. (2020)

This study examined mountain glacial melt in Asia using a new resource, stereo imagery from commercial satellites, with the aim of improving quantitative estimates of melt and contribution to global sea level rise. Like other articles in this section, the abstract starts with a brief Introductory Funnel, followed by Evidence and Reasoning before coming to the main Claim.

Title: “A systematic, regional assessment of high mountain Asia glacier mass balance”

Abstract:

“High-mountain Asia (HMA) constitutes the largest glacierized region outside of the Earth's polar regions. Although available observations are limited, long-term records indicate sustained HMA glacier mass loss since ~1850, with accelerated loss in recent decades. Recent satellite data capture the spatial variability of this mass loss, but spatial resolution is coarse and some estimates for regional and HMA-wide mass loss disagree. To address these issues, we generated 5,797 high-resolution digital elevation models (DEMs) from available sub-meter commercial stereo imagery (DigitalGlobe WorldView-1/2/3 and GeoEye-1) acquired over HMA glaciers from 2007 to 2018 (primarily 2013–2017). We also reprocessed 28,278 ASTER DEMs over HMA from 2000 to 2018. We combined these observations to generate robust elevation change trend maps and geodetic mass balance estimates for 99% of HMA glaciers between 2000 and 2018. We estimate total HMA glacier mass change of $-19.0 \pm 2.5 \text{ Gt yr}^{-1}$ ($-0.19 \pm 0.03 \text{ m w.e. yr}^{-1}$). We document the spatial pattern of HMA glacier mass change with unprecedented detail, and present aggregated estimates for HMA glacierized sub-regions and hydrologic basins. Our results offer improved estimates for the HMA contribution to global sea level rise in recent decades with total cumulative sea-level rise contribution of ~0.7 mm from exorheic basins between 2000 and 2018. We estimate that the range of excess glacier meltwater runoff due to negative glacier mass balance in each basin constitutes ~12–53% of the total basin-specific glacier meltwater runoff. These results can be used for calibration and validation of glacier mass balance models, satellite gravimetry observations, and hydrologic models needed for present and future water resource management.”

Let's look at the Introductory Funnel presented in the first part of the abstract:

- **“High-mountain Asia (HMA) constitutes the largest glacierized region outside of the Earth's polar regions.”**

The first sentence establishes the broad context (glaciers) and the scope of the study (high-mountain Asian glaciers), noting the relevance of this region as “the largest glacierized region outside of the Earth’s polar regions.”

- **“Although available observations are limited, long-term records indicate sustained HMA glacier mass loss since ~1850, with accelerated loss in recent decades.”**

The second sentence accomplishes several things: it defines the key feature under investigation (HMA glacier mass loss) and states the problem (sustained and recently accelerated mass loss). It also introduces a research gap based on the limited observations currently available for this region.

- **“Recent satellite data capture the spatial variability of this mass loss, but spatial resolution is coarse and some estimates for regional and HMA-wide mass loss disagree.”**

The third sentence clearly defines the main research gaps: the coarse resolution of current satellite data and disagreements between mass loss estimates made regionally and across the HMA area.

What follows in the abstract is a series of Evidence and Reasoning statements which address these research gaps, most notably resolution (“sub-meter commercial stereo imagery”) and the reanalysis of data to expand the observational dataset. The authors use these to support several Claims:

- We estimate total HMA glacier mass change of $-19.0 \pm 2.5 \text{ Gt yr}^{-1}$ ($-0.19 \pm 0.03 \text{ m w.e. yr}^{-1}$).

This statement provides a quantitative determination of HMA mass loss that addresses the “limited available observations,” and is also made possible because of the improved resolution of the data.

- Our results offer improved estimates for the HMA contribution to global sea level rise in recent decades with total cumulative sea-level rise contribution of $\sim 0.7 \text{ mm}$ from exorheic basins between 2000 and 2018.

This statement further expands on the importance of HMA glacial melt (“exorheic” is drainage from land to ocean) in contributing to global sea level rise, made possible by the more precise measurements made by the new dataset.

- We estimate that the range of excess glacier meltwater runoff due to negative glacier mass balance in each basin constitutes $\sim 12\text{--}53\%$ of the total basin-specific glacier meltwater runoff.

This last Claim narrows down the specific contribution of glacial melt from other runoff sources, again made possible by the improved resolution.

Taken together, the authors are able to use a new class of observations to address key shortcomings in measuring glacial melt in high Asian mountains and quantify the contribution of this melt to global sea level rise.

SS 7.5 Physics

Example: Rüter et al. (2010)

This article presents a new approach to studying quantum systems using optics, where theoretical aspects of parity-time (PT) symmetry can be directly explored through analog experimental systems. As this article bridges two very disparate fields (quantum theory and optics), the abstract provides an extensive and well-cited Introductory Funnel to motivate the study.

Title: “Observation of parity–time symmetry in optics”

Abstract:

“One of the fundamental axioms of quantum mechanics is associated with the Hermiticity of physical observables (ref. 1). In the case of the Hamiltonian operator, this requirement not only implies real eigenenergies but also guarantees probability conservation. Interestingly, a wide class of non-Hermitian Hamiltonians can still show entirely real spectra. Among these are Hamiltonians respecting parity–time (*PT*) symmetry (refs. 2,3,4,5,6,7). Even though the Hermiticity of quantum observables was never in doubt, such concepts have motivated discussions on several fronts in physics, including quantum field theories (ref. 8), non-Hermitian Anderson models (ref. 9) and open quantum systems (refs. 10,11), to mention a few. Although the impact of *PT* symmetry in these fields is still debated, it has been recently realized that optics can provide a fertile ground where *PT*-related notions can be implemented and experimentally investigated (refs. 12,13,14,15). In this letter we report the first observation of the behaviour of a *PT* optical coupled system that judiciously involves a complex index potential. We observe both spontaneous *PT* symmetry breaking and power oscillations that violate left–right symmetry. Our results may pave the way towards a new class of *PT*-synthetic materials with intriguing and unexpected properties that rely on non-reciprocal light propagation and tailored transverse energy flow.”

Let’s break down the first three-quarters of this abstract using the Introductory Funnel model:

- **“One of the fundamental axioms of quantum mechanics is associated with the Hermiticity of physical observables (ref. 1).”**

The first sentence establishes the broad context (quantum mechanics) and the key theme (Hermiticity of physical observables, a requirement for real observables in quantum theory).

- **“In the case of the Hamiltonian operator, this requirement not only implies real eigenenergies but also guarantees probability conservation.”**

The next sentence adds more clarity to the them by providing an example with the Hamiltonian operator, which measures the energy of quantum system and ensures that quantum wave functions have finite probability.

- **“Interestingly, a wide class of non-Hermitian Hamiltonians can still show entirely real spectra. Among these are Hamiltonians respecting parity–time (*PT*) symmetry (refs. 2,3,4,5,6,7). Even though the Hermiticity of quantum observables was never in doubt, such concepts have motivated discussions on several fronts in physics, including quantum field theories (ref. 8), non-Hermitian Anderson models (ref. 9) and open quantum systems (refs. 10,11), to mention a few.”**

The next three sentences introduce a new theme, the existence of non-Hermitian Hamiltonians that nevertheless lead to real observables. While the existence of these Hamiltonians, particularly those that respect parity-time symmetry (reflection in space and time), does not necessary violate the fundamental axiom that physical observables are Hermitian, they have been the source of theoretical investigations of quantum systems, indicating their importance.

- **“Although the impact of *PT* symmetry in these fields is still debated, it has been recently realized that optics can provide a fertile ground where *PT*-related notions can be implemented and experimentally investigated (refs. 12,13,14,15).”**

This next sentence reveals the key research gap that motivates this study, namely the ongoing theoretical debate of the impact of *PT* symmetry in these quantum systems; and the possible resolution of this gap through the use of optics as an experimental analog.

- “In this letter we report the first observation of the behaviour of a *PT* optical coupled system that judiciously involves a complex index potential.”

Finally, we get to the main Discovery Claim, demonstrating measurement of an optics system that models a non-Hermitian (“complex”) Hamiltonian (“potential”).

Discovery papers such as this often have an extensive Introductory Funnel in the abstract or introduction section, as both the motivation for the search and relevance of the discovery need to be properly contextualized. Here, Rüter et al. (2010) must also provide the connection between two seemingly distinct physical systems: theoretical quantum systems with non-Hermitian Hamiltonians and experimental optics systems. By providing an extensive discussion of the Context in the abstract, the reader can better understand both the motivation for the study and its broader Implications across physical science fields.

Student Supplement (SS) for Chapter 9: Using CERIC to Compare Articles

Scientific progress unfolds as a conversation: one study builds upon another, refines an idea, or challenges an existing framework. Comparing research articles is a way to trace this intellectual dialogue, revealing how studies complement, refine, or contradict one another. Whether assessing competing theories, different methodologies, or conflicting results, effective comparison deepens our understanding of a research field and sharpens critical thinking.

SS 9.1 Establishing a Comparative Framework

A well-structured comparison begins with defining key evaluation criteria. These include:

- **Research questions and hypotheses:** Do the studies investigate the same scientific question, or do they approach related but distinct problems? If different, how do their scopes intersect?
- **Methodological approaches:** How do study designs, sample sizes, data collection techniques, and statistical analyses compare? Which approach provides stronger empirical support based on rigor and reproducibility?
- **Key findings and Evidence strength:** Do results align across studies, or do discrepancies suggest alternative explanations? What levels of statistical significance and reproducibility are reported?
- **Interpretation and theoretical contributions:** Do the studies reinforce, refine, or challenge existing theories?
- **Limitations and biases:** What constraints do the authors acknowledge? Are there methodological or funding-related factors that might influence interpretations?
- **Broader Implications:** How do the conclusions shape policy, applications, or future research?

For example, in astronomy, studies of exoplanet atmospheres often compare spectral data collected by different telescopes under varying observational conditions. One study may infer the presence of water vapor based on near-infrared absorption, while another might detect a different chemical fingerprint at longer mid-infrared wavelengths. Comparing these studies involves evaluating their findings and considering how differences in instrument sensitivity, atmospheric models, and statistical approaches shape their conclusions.

SS 9.2 Evaluating Methodological Rigor and Reliability

Methodological rigor influences the credibility and impact of scientific findings. A comparative analysis assesses how studies differ in:

- **Study design and controls:** Does one study implement more robust experimental controls or use a more comprehensive dataset?

- **Sample size and statistical power:** Do the findings hold across large datasets, or are conclusions based on small samples with marginal significance?
- **Data sources and transparency:** Do the authors provide raw data and detailed methods, ensuring reproducibility?
- **External validity and generalizability:** Are the findings applicable beyond the experimental conditions, or do they require field studies for confirmation?

For instance, in biomedical research, a laboratory study may demonstrate that a CRISPR-based gene therapy restores retinal function in mice (Qin et al., 2023). However, a comparative study might highlight a clinical trial testing a similar therapy in human patients, raising questions about scalability, immune responses, and long-term efficacy. Evaluating these methodological differences provides insight into how findings translate from controlled experiments to real-world applications.

SS 9.3 Identifying Trends and Contradictions Across Studies

Patterns in the literature reveal whether a scientific field is moving toward consensus, divergence, or transformation.

- **Converging Evidence:** If multiple studies reach similar conclusions using distinct methodologies, the result gains credibility.
- **Conflicting Results:** Discrepancies may arise from methodological differences, sample variations, or contextual factors—each requiring careful examination.
- **Evolution of the Field:** Comparing older and newer studies highlights how research progresses. Have recent studies refined experimental techniques, introduced new theoretical models, or corrected previous assumptions?

The following example illustrates the process of comparing trends in a related pair of climate science studies investigating coral resilience.

Example: Comparing Coral Acclimatization and Genetic Modification

Consider two studies investigating how corals might survive rising ocean temperatures:

- **Study A:** Tests acclimatization by exposing corals to gradually increasing temperatures, showing that heat tolerance improves over generations.
- **Study B:** Uses genetic modification to introduce heat-tolerance genes, leading to immediate but possibly short-term resilience.

A comparative analysis would examine:

- **Effectiveness:** Does acclimatization provide similar resilience benefits as genetic intervention?
- **Scalability and Ethical Considerations:** Which approach is more feasible for large-scale reef conservation?

- **Long-Term Stability:** Will acclimatization effects persist across generations, or does genetic modification offer a more stable adaptation strategy?

A well-structured comparison clarifies not just which method is effective but also how each aligns with ecological, ethical, and policy considerations.

SS 9.4 Synthesizing Comparisons for Academic and Professional Discussions

A strong comparative analysis moves beyond listing similarities and differences to drawing meaningful conclusions. This comparative synthesis should:

- Highlight key points of agreement and divergence.
- Assess the relative strengths and limitations of each approach.
- Propose future research directions based on collective Evidence.

For example, a discussion comparing mRNA vaccines and traditional inactivated-virus vaccines would assess effectiveness, durability of the immune response, and scalability. While both strategies provide protection, mRNA vaccines offer faster adaptation to new variants, whereas inactivated virus vaccines may be more accessible in lower-resource settings. A balanced synthesis would outline trade-offs and suggest next steps for optimizing vaccine deployment.

Student Supplement for Chapter 10: Critique

Critique of the scientific literature is an essential component of the scientific process, underpinning the process of peer review. Critique extends our reading from understanding to critical analysis of a study's methods, measurements, and findings, touching each of the CERIC elements. The following is a summary of how to approach article critique from the perspective of a peer reviewer using CERIC as an organizing guide. Further detail and examples of CERIC critique are provided in Chapter 10.

SS 10.1 Examine the Claim

Start by identifying the main Claim or research question of the manuscript. Is it clearly stated in the introduction and abstract? During the first read, a reviewer must ensure that they understand *what* the authors are setting out to prove or investigate. Then, critically evaluate this Claim: *Is it significant and novel? Is it testable or answerable with the presented approach?* As a reviewer, ask yourself whether the Claim is too broad or appropriately focused. If the Claim is muddled or not explicitly stated, that's feedback to give. If the Claim is clear, consider whether it's interesting and valuable to the field (editors often rely on reviewers to judge the importance of research findings). This step sets the stage for the rest of the review, since all other comments will relate to how well the Claim is supported by Evidence and Reasoning.

SS 10.2 Scrutinize the Evidence

Next, delve into the methodology and results – the Evidence of an article. A conscientious reviewer will check the experimental design, data collection, and analysis with a fine-tooth comb. Key questions include: *Are the methods appropriate to gather the necessary Evidence for the Claim? Are sample sizes sufficient and are controls in place? Do the figures and tables accurately reflect the data and are they clear?* Also look for errors or biases: e.g., was a proper statistical test used, or could the results be attributable to chance (Type I error) or missed due to low power (Type II error)? Reviewers often replicate, at least conceptually, parts of the analysis – recalculating an easy statistic or ensuring that reported values make sense. If something in the data looks fishy or inconsistent, the reviewer should point it out.

It is also important to consider whether the Evidence is complete: *Do the authors address all relevant results, including any that contradict their hypothesis?* For instance, if certain data points were excluded or an outlier was ignored without justification, a reviewer should flag that. In essence, the reviewer's job is to validate that the Evidence presented is reliable, sufficient, and directly relevant to the Claim (Mohty & Melo, 2025). Any shortcomings (an uncontrolled variable, a possible confounder, or unclear data presentation) should be noted, along with suggestions for improvement.

SS 10.3 Evaluate the Reasoning

After assessing that the data are sound, the reviewer examines how the authors connect the Evidence to their Claim. This means reading the discussion/analysis sections carefully. Check whether the authors' interpretations are logical and whether they address alternative explanations. A peer reviewer using CERIC will ask: *“Do the Reasoning steps support the Claim given the Evidence?”* If the authors make a big inference that isn't directly supported, that's a point for critique. Look out for classic Reasoning fallacies – for example, are the authors assuming causation from a correlation? Are they ignoring plausible opposing interpretations of the data? Many journals expect reviewers to comment on whether the conclusions are justified (Kelly et al., 2014; Korner, 2008). It is helpful to enumerate specific instances. For example, *“In paragraph 3 of the discussion, the authors conclude X, but this doesn't necessarily follow because Y could also explain the result.”* The reviewer should also check consistency: the findings reported in the abstract should match those in the main body of the paper

In practice, a good peer review often includes a statement like *“The data support the authors' primary claim overall, except for [some caveat] where the logic was not entirely convincing.”* By systematically focusing on Reasoning, the reviewer helps ensure the authors don't make Claims beyond what their Evidence allows.

SS 10.4 Consider the Implications

A peer review should also address the study's significance and Implications. A peer reviewer using CERIC will ask: *Do the authors adequately discuss the Implications of the findings for the field? Do they avoid overstating the Implications?* In applying CERIC, one should ensure that the implications section of the paper (often the last part of the discussion or the conclusions section) is sound. Are the Implications supported by the findings? Are speculative Implications appropriately qualified with phrases like *“if confirmed, these results suggest...”* rather than absolute statements? A critical peer review might include a statement like: *“The authors need to temper their conclusions about clinical relevance; the sample is small, so broad public health implications are premature.”* By doing so, the reviewer helps maintain a realistic interpretation of the study's importance.

The CERIC framework also reminds the reviewer to ensure that the authors have framed their conclusions responsibly within the broader context. If the authors claim their discovery “revolutionizes” a field or has a massive real-world impact, the reviewer should assess if that's an exaggeration. Conversely, if the authors haven't mentioned an important Implication or limitation, the reviewer might suggest adding it. Ethical issues also fall under this category. For instance, if a study's results could be misused or have ethical ramifications (for example, research on dual-use technology or sensitive social topics), the reviewer can recommend that the authors comment on them.

SS 10.5 Review the Context

Finally, a CERIC-guided peer review examines the research Context. This involves checking the introduction section and references to ensure the authors are building their study on solid ground. The reviewer should verify that the authors cited relevant previous studies, especially any seminal works or recent key findings relevant to the topic of their study. If key references are missing, the reviewer can point them out, for example: *“Prior work by Smith et al. (2018) is highly relevant and should be discussed.”* Review of Context also includes evaluating whether the study’s design and scope are appropriate given the current state of knowledge. Does the paper address a real gap in the literature, or does it seem to reinvent the wheel? If a reviewer knows that similar data have been published before, they should alert the editor and the authors.

Additionally, the reviewer can judge whether the background provided is accurate and sufficient for readers. Sometimes authors may misrepresent prior studies intentionally or unintentionally; a vigilant reviewer will catch discrepancies and note in their review *“The authors claim no one has studied this before, but a 2015 study by Smith et al. report a comparable study that should be discussed in the article.”* Comments in this vein might include suggestions to compare the results with those of other studies or to discuss conflicting evidence from the literature. Ensuring proper context also means confirming the work is relevant to the journal’s audience and fits the broader scientific dialogue. In sum, by evaluating Context the reviewer helps uphold scholarly continuity, correctness, and completeness.

Student Supplement (SS) for Chapter 12: Literature Reviews

This supplement explores how student researchers can transition from applying the CERIC method at the level of individual articles to synthesizing broader scientific narratives across multiple studies.

SS 12.1 Synthesize Multiple Studies

By integrating the elements of Claim, Evidence, Reasoning, Implications, and Context, researchers can develop cohesive understandings of complex scientific phenomena. Here, we provide clear, illustrative examples that demonstrate how the synthesis of CERIC elements, individually or in combination, can reveal overarching patterns and insights from climate science research.

SS 12.1.1 Example: Synthesizing Multiple Claims

Synthesizing Claims from recent climate science research reveals that (1) global temperatures are rising significantly due to human activity, particularly fossil fuel combustion; (2) ecosystems are shifting geographically in response to climate stressors; (3) ocean acidification impacts marine biodiversity by disrupting reproductive and growth processes; (4) deforestation contributes directly to biodiversity loss and indirectly to climate change through carbon release; and (5) climate-induced extreme weather events are becoming more frequent and intense (IPCC, 2021; Davis et al., 2005; Barbour & Gilbert, 2021; Franks & Hoffman, 2012).

SS 12.1.2 Example: Synthesizing Evidence Across Studies

Integrating Evidence across ecological and environmental studies provides strong empirical support for climate-related claims, such as: (1) ice-core data demonstrating unprecedented increases in atmospheric carbon dioxide; (2) long-term biodiversity monitoring revealing species migrations toward poles and higher elevations; (3) experimental studies showing decreased calcification in marine organisms under acidic conditions; (4) satellite imagery documenting accelerated deforestation rates and resultant habitat loss; and (5) meteorological records verifying rising incidences of floods, droughts, and storms (IPCC, 2021; Davis et al., 2005; Hoegh-Guldberg et al., 2007; Barbour & Gilbert, 2021).

SS 12.1.3 Example: Synthesizing Lines of Reasoning

The Reasoning synthesized across multiple studies indicates: (1) fossil fuel combustion directly elevates greenhouse gas concentrations, trapping heat within Earth's atmosphere; (2) shifts in species distributions occur as organisms seek suitable climates, supporting ecological niche theory; (3) increased carbon dioxide absorption by oceans alters carbonate chemistry, impeding marine calcification processes; (4) forest clearing disrupts carbon storage capacity, exacerbating atmospheric carbon accumulation; and (5) enhanced atmospheric instability due to warming intensifies weather events through increased

moisture retention and heat exchange dynamics (Hoegh-Guldberg et al., 2007; Franks & Hoffman, 2012; IPCC, 2021).

SS 12.1.4 Example: Synthesizing Research Implications

Synthesizing Implications from recent climate and ecological research underscores: (1) a pressing need for mitigation policies to curb greenhouse gas emissions; (2) prioritizing conservation efforts to protect climate-sensitive biodiversity hotspots; (3) adaptation strategies to support marine ecosystems under increased ocean acidification; (4) reforestation and afforestation projects to stabilize carbon sequestration; and (5) investments in resilient infrastructure to withstand increasingly frequent extreme weather events (IPCC, 2021; Barbour & Gilbert, 2021; Davis et al., 2005).

SS 12.1.5 Example: Synthesizing Contextual Gaps

Synthesizing Contexts from multiple studies highlights current or persistent research gaps: (1) incomplete understanding of how specific ecosystems will adapt to rapidly changing climates; (2) uncertainty surrounding the genetic mechanisms enabling rapid evolutionary responses; (3) limited data on long-term effects of ocean acidification on broader marine food webs; (4) insufficient integration of socio-economic impacts of deforestation in global climate policy discussions; and (5) knowledge gaps about precise regional climate predictions hindering local adaptation efforts (Franks & Hoffman, 2012; Hoegh-Guldberg et al., 2007; Barbour & Gilbert, 2021).

SS 12.1.6 Example: Synthesizing a C-E-R Core Argument

Articulating the core C-E-R argument across multiple studies provides an efficient way to synthesize their interconnected components. Drawing on the above examples, we can state the following C-E-R argument from the existing literature: Climate change significantly affects ecological stability and biodiversity through mechanisms such as altered temperature regimes, ocean acidification, and habitat fragmentation (Claim). Empirical data support these findings through longitudinal shifts in biodiversity, experimental observations of reduced calcification, and documented increases in extreme weather events (Evidence). The logical arguments supporting these effects involves interconnected processes such as atmospheric carbon accumulation from fossil fuels, ecological displacement due to shifting niches, and chemical disruptions to marine life cycles (Reasoning) (IPCC, 2021; Davis et al., 2005; Hoegh-Guldberg et al., 2007).

Comparing individual CERIC elements and CER core arguments across studies provides a foundation for moving from *analysis* to *synthesis*. However, the real strength of this approach lies in its adaptability. Depending on research goals or the complexity of scientific narratives, researchers can selectively combine these elements to reveal nuanced trends, logical connections, and compelling opportunities for future investigation. Ultimately, the skillful

integration of Claims, Evidence, Reasoning, Implications, and Context informs a broader and richer understanding and sets a clear, strategic path forward for future research.

SS 12.2 Literature Reviews in Theses and Dissertations

Thesis and dissertation literature reviews explicitly link the existing literature to the research questions or hypotheses of a graduate-level investigation. In such work, the graduate researcher must make an original contribution of knowledge to the field. Justifying the originality and motivation of the work usually begins with a literature review, which may be revisited many times throughout the study.

Let's look at two examples of how graduate researchers identified scientifically relevant gaps through literature-driven analysis.

SS 12.2.1 Example: Molecular and Cell Biology Thesis

A doctoral dissertation by Catia Caetano published by University College London examined the regulation of gene expression during the G1/S transition in the cell cycle and its impact on genome stability in *Saccharomyces cerevisiae* (Caetano, 2012). The author identified gaps in the understanding of DNA repair pathways:

"Activation of the DNA replication checkpoint impacts on cell cycle progression by eliciting many responses, including the initiation of transcription. However, how this effect is mediated in budding yeast is not entirely clear" (p. 86).

The author proposed a specific experimental hypothesis to address the gap:

"Data from fission yeast research has identified a mechanism whereby the spMBF-dependent G1/S transcripts are de-repressed in response to activation of the DNA replication checkpoint...() I hypothesise that a similar mechanism may be present in budding yeast" (p. 86).

The author also proposed examining the involvement of specific co-repressors and kinases in this process. These excerpts illustrate how graduate researchers using the existing research to identify a research gap and propose a way to address it.

SS 12.2.2 Example: Neuroscience Thesis

A doctoral dissertation by Christine Denny published by Columbia University on neuronal signaling synthesized Evidence regarding the roles of various neurotransmitters in memory processes, examined experimental inconsistencies, and connected those findings to hypotheses concerning cognitive disorders such as Alzheimer's disease and depression (Denny, 2012). The literature review identified a significant gap in understanding the functional contributions of adult-born neurons to hippocampus-dependent behaviors:

"Despite the consensus that adult-born neurons are functionally significant, there is little or no agreement on which behavioral tasks are sensitive to the disruption of adult neurogenesis and on which underlying psychological processes are altered when adult neurogenesis is arrested" (p. 28)

The author then proposed investigations using activity-dependent neural tagging techniques in mice to explore the neural mechanisms underlying learning and memory:

"We sought to determine whether different CFC methods vary in their sensitivity to the arrest of adult neurogenesis. We arrested adult neurogenesis using low-dose, targeted x-irradiation, and then assayed the effects of irradiation in conditioning procedures that varied in the use of a discrete conditioned stimulus, the number of trials administered, and the final level of conditioning produced" (p. 30)

The author also provides Reasoning for the choice of investigative approach:

"Irradiation was selected because it is a common method of arresting neurogenesis employed by several laboratories, produces a complete and permanent arrest of adult neurogenesis, and leaves other neurogenic niches intact when applied focally (Meshi et al., 2006; Santarelli et al., 2003)" (p. 30).

These two examples demonstrate how literature-driven analysis enables graduate researchers to strategically identify scientific gaps and thoughtfully propose targeted experiments to advance understanding in their respective fields.

SS 12.3 Stand-Alone Literature Reviews

Stand-alone literature reviews offer an extensive synthesis of existing Claims, Evidence, and Reasoning arguments in a field of study, highlighting research trends, gaps, and future directions. Let's examine two case studies of stand-alone review articles and apply CERIC as a framework to understand the main arguments in each case.

SS 12.3.1 Case Study 1: CRISPR-Cas9 Technology Developments

A comprehensive review by Doudna and Charpentier (2014) synthesized more than 150 articles (indicated as numbers in parentheses) on the groundbreaking potential of CRISPR-Cas9 for precise genome editing. The challenge of such an extensive review is to find the CERIC elements in synthesized clusters of articles.

Here's an example of a synthesis paragraph integrating ten prior Claims:

“CRISPRs were later detected in numerous bacteria and archaea (35), and predictions were made about their possible roles in DNA repair or gene regulation (36, 37). A key insight came in 2005 with the observation that many spacer sequences within CRISPRs derive from plasmid and viral origins (38–40). Together with the finding that CRISPR loci are transcribed (41) and the observation that cas (CRISPR-associated) genes encode proteins with putative nuclease and helicase domains (38, 40, 42, 43), it was proposed that CRISPR-Cas is an adaptive defense system that might use antisense RNAs as memory signatures of past invasions (44)” (p. 1258096-1).

The authors meticulously analyzed multiple lines of Evidence on efficiency and off-target effects, employing detailed molecular Reasoning to explain how CRISPR technology achieves its renowned accuracy. Here’s an example of a synthesis paragraph integrating three lines of Evidence:

“Results showed that in both human embryonic kidney (HEK293) cells (83) and mouse embryonic stem cells (mESCs) (84), a catalytically inactive version of Cas9 bound to many more sites than those matching the sequence of the sgRNA used in each case. Such off-target interactions with DNA, typically at sites bearing a PAM and partially complementary to the guide RNA sequence, are consistent with established modes of DNA interrogation by Cas9 (80)” (p. 1258096-2).

The authors provide Reasoning, such as consistency with prior theory, results, limitations, and alternative hypotheses. Here’s an example of a synthesis paragraph integrating three lines of Reasoning (prior results, limitation, and alternate hypothesis):

“These observations are consistent with results obtained for Cas9–guide RNA complexes in single-molecule experiments (80). Furthermore, Cas9 binding events occur more densely in areas of open chromatin as compared to regions of compact, transcriptionally inactive chromatin. However, because the method involves cross-linking cells for ~10 min before quenching the reaction, transient and long-lived binding interactions cannot be distinguished. It is possible that many of the apparent off-target DNA interactions in fact reflect brief encounters that would not normally trigger strand invasion by the guide RNA” (p. 1258096-2).

Crucially, the authors articulated a vision for future research, effectively framing the Implications as priorities. Here’s an example of a synthesis paragraph identifying three fields as major directions (i.e., biology, biotechnology, and therapeutics):

“These results highlight a new era in which genomic manipulation is no longer a bottleneck to experiments, paving the way toward fundamental discoveries in biology, with applications in all branches of biotechnology, as well as strategies for human therapeutics” (p. 1258096-1, abstract).

The review also detailed specific examples of Implications. Here's an example of a synthesis paragraph integrating sources that implicate gene therapy as a future line of research:

“Furthermore, specific methods for delivering Cas9 and its guide RNA to cells and tissues should benefit the field of human gene therapy. For example, recent experiments confirmed that the Cas9 protein-RNA complex can be introduced directly into cells using nucleofection or cell-penetrating peptides to enable rapid and timed editing (89, 152), and transgenic organisms that express Cas9 from inducible promoters are being tested. An exciting harbinger of future research in this area is the recent demonstration that Cas9–guide RNA complexes, when injected into adult mice, provided sufficient editing in the liver to alleviate a genetic disorder (153)” (p. 1258096-7).

Additionally, the review explored the broad Implications for bioethics, including potential societal and ethical concerns related to genetic modifications. Here's an example of a synthesis paragraph integrating several ethical concerns:

“The era of straightforward genome editing raises ethical questions that will need to be addressed by scientists and society at large. How can we use this powerful tool in such a way as to ensure maximum benefit while minimizing risks? It will be imperative that nonscientists understand the basics of this technology sufficiently well to facilitate rational public discourse. Regulatory agencies will also need to consider how best to foster responsible use of CRISPR-Cas9 technology without inhibiting appropriate research and development” (p. 1258096-7).

Context synthesis is abundant in this review article. Here's an example of a synthesis paragraph integrating several historical developments in DNA-targeted cleavage:

“Building on the original description of triple helix formation by Rich and colleagues in the late 1950s (10, 11), oligonucleotides coupled to chemical cleavage or cross-linking reagents such as bleomycin and psoralen were shown to be useful for site-specific chromosome modification in yeast and mammalian cells (12–17)” (p. 1258096-1).

The authors identify several historical gaps in the research that drove it forward, for example:

“... difficulties of protein design, synthesis, and validation remained a barrier to widespread adoption of these engineered nucleases for routine use” (p. 1258096-1).

Finally, the authors provide a simple statement of the provided Context, and a motivation for the review in the scope of new developments:

“Below we discuss the history and biology of CRISPR systems, describe the molecular mechanisms underlying genome editing by Cas9, and review the rapid advances in applications of this technology since its initial publication in 2012” (p. 1258096-1).

SS 12.3.2 Case Study 2: Global Biodiversity Assessments

The comprehensive Global Assessment Report on Biodiversity and Ecosystem Services by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES, 2019) integrated more than 150 source articles on global biodiversity loss. The report is incredibly dense and daunting at nearly 1,200 pages. Nonetheless, applying CERIC as a framework can help readers grasp the main points, while acknowledging the challenge in processing this much information in a single pass.

The IPBES (2019) report systematically reviewed dozens of Claims about biodiversity and ecological services across varied ecosystems. Here’s an example of a synthesis paragraph integrating several findings regarding the primary drivers affecting biodiversity, consolidating them into a comprehensive overview:

“The direct drivers of change in nature with the largest global impact have been (starting with those with most impact): changes in land and sea use; direct exploitation of organisms; climate change; pollution; and invasion of alien species” (p.XVI)

The report critically evaluates many diverse lines of ecological Evidence. Here’s an example of a synthesis paragraph integrating agricultural output, fisheries data, and bioenergy statistics, to illustrate the correlation between human activity and environmental impact:

“Since 1970, trends in agricultural production, fish harvest, bioenergy production and harvest of materials have increased, responding to population growth, rising demand and technological development, although negative impacts on nature and its contributions to people have increased” (p. XV).

The authors extensively discuss the key Reasoning elements in the literature, including complex, interconnected drivers such as habitat destruction, climate change, invasive species, and pollution. Here’s one example of a synthesis paragraph integrating counterfactuals, which are less common to see in reviews:

“This disconnect between progress in responses and increases in drivers of change in nature and NCP requires consideration. There is not a simple linear relationship, owing to several reasons. First, from a small set of counterfactual studies and other assessments (e.g., Geldmann et al., 2013; Hoffmann et al., 2010, 2015; Jones et al., 2016; Waldron et al., 2017), trends in drivers and the state of nature would be worse

without the conservation responses that have been implemented (section 3.2)” (p. 507)

By clearly outlining urgent Implications for conservation policy and management, biodiversity threats and the significant gaps in current knowledge, the report lays out an agenda for future research:

“At the higher level examining the relationship between nature, NCP and education, there is growing work and evidence on the role of access to nature and urban green space for achieving education outcomes (Mocior & Kruse, 2016) as well as in aspects relevant to education including cognitive function and mental health (e.g., Brattman et al., 2012, 2015). This is a promising area of future research, especially considering the knock-on effect of education on achievement of other SDGs (Figure S3.3)” (p. 468).

Context synthesis is abundant in this review article. Here’s an example of a synthesis paragraph referencing species extinctions since the 16th century, linking past human activities to current biodiversity threats:

"Human actions have already driven at least 680 vertebrate species to extinction since 1500, and more than 40% of amphibian species, almost 33% of reef-forming corals, and more than a third of all marine mammals are threatened" (p. XXVIII)

The authors identify dozens of gaps concerning knowledge in multiple sectors, for example:

"Knowledge gaps in data, indicators, inventories, [and] scenarios" sectors (p. LIX)

Finally, the authors describe the Context that motivated the report, highlighting the inadequacy of current efforts and the need for a comprehensive assessment to inform more effective strategies:

“Current evaluations (chapters 2, 3) and most future scenarios (chapter 4) show that goals for conserving and sustainably using nature and achieving sustainability cannot be met by current trajectories, and goals for 2030 and beyond, the 2020 Aichi Biodiversity Targets, and Paris Agreement on Climate Change may only be achieved through transformative changes across economic, social, political and technological factors" (p. 772)

These examples illustrate how applying CERIC as a framework can reveal essential connections among complex review articles. Evaluating each CERIC element provides a clear, structured pathway to understanding the depth and coherence of the arguments presented. It is worth the effort, especially in well-written review articles that accurately and comprehensively synthesize current scientific knowledge. Working through these kinds of

reviews is a way for learners to undertake a deeper dive into a particular field, and for experts to guide discourse on future research priorities, policies, and ethical conversations in science.

SS 12.4 Historical Development of a Field

Another engaging strategy for structuring a literature review is narrating a topic's historical development. This chronological approach presents the literature in roughly the order it was published or discovered, vividly capturing the evolution of ideas and breakthroughs. Such a narrative is particularly valuable in fields characterized by distinct phases or transformative paradigm shifts. Rather than merely summarizing research year by year, an effective historical review highlights significant trends, critical turning points, and influential eras, analyzing their contributions to contemporary understanding (Galvan & Galvan, 2017).

The CERIC framework elegantly guides this historical narrative:

- **Claim** (for each era or phase): A clear review divides the timeline into meaningful segments—for example, “Early Research (1990–2000),” “Mid-Development (2000–2010),” and “Recent Advances (2010–present).” The review clearly articulates the dominant understanding or prevalent approaches within each period, such as stating, “During the 1990s, scientists believed high-temperature superconductivity was limited exclusively to copper-based compounds.”
- **Evidence:** Each Claim is substantiated by highlighting key studies or groundbreaking papers representative of the era. The relevant Evidence could include influential experiments, significant theoretical models, or pivotal research outcomes.
- **Reasoning:** The narrative explains why and how the field has evolved. It examines technological innovations (such as advances in research instruments), shifts in theoretical frameworks (such as new hypotheses or models), or external drivers (such as shifts in funding or societal demands) that influenced the direction of research. Reasoning clearly connects the Evidence to the Claim, showcasing a cause-and-effect progression in knowledge development. For example, “In the early 2000s, powerful new telescopes enabled astronomers to collect more detailed data, prompting the revision of existing theories.”
- **Implications:** At crucial turning points, the review examines how findings from each period influenced subsequent research, resolved debates, or posed new questions. For instance, a critical discovery might have resolved long-standing controversies or revealed anomalies that puzzled researchers, setting the stage for future inquiry.
- **Context:** Historical context enriches the narrative by highlighting technological capabilities, prevailing scientific paradigms, and concurrent global or societal events. Contextual insights help readers grasp why particular questions emerged and received attention in specific periods.

Let’s take a deeper dive into a historical narrative literature review example from computer science.

SS 12.4.1 Example: Computer Science

Consider the literature review by Haenlein and Kaplan (2019), which traces the development of artificial intelligence (AI) from the 1950s to the present. The review employs a structured series of Claim-Evidence-Reasoning argument cores to articulate distinct phases in AI's evolution. Implications and Context are woven throughout.

The narrative begins with a single Claim: AI research originated in the mid-20th century and was initially driven by symbolic approaches and strong optimism about rapid progress (Haenlein & Kaplan, 2019).

This assertion of AI's spring or birth is supported by Evidence, including the seminal Dartmouth Summer Research Project on Artificial Intelligence in 1956, led by pioneers like John McCarthy and Marvin Minsky, and the development of early symbolic AI programs such as ELIZA and the General Problem Solver. Reasoning highlights that limited computational resources and the philosophical belief that intelligence could be formally described led researchers naturally toward symbolic reasoning methods and expert systems.

Transitioning into the 1970s and early 1980s, the review next Claims a period known as the "AI Winter," characterized by significant funding cuts and skepticism regarding AI's potential. This claim is supported by Evidence such as the critical Lighthill Report of 1973, which cast doubt on the feasibility of achieving general AI, leading governments to substantially reduce their support for AI research. Reasoning holds that the initial enthusiasm, based on overly optimistic predictions, was not matched by the capabilities of early expert systems, leading to widespread disillusionment.

In the late 1980s and throughout the 1990s, the review Claims a revival of AI—an "AI Fall" or harvesting period—marked by practical successes in expert systems and significant victories in highly structured games. This is supported by Evidence, including IBM's Deep Blue victory over world chess champion Gary Kasparov in 1997. Reasoning connects this development to advances in computational capacity and the refinement of expert systems, demonstrating that AI could surpass human capabilities in specific, clearly defined domains.

Finally, moving into the modern era (2010s–present), the review Claims the current age of "AI Summer" is dominated by deep learning and neural networks, signifying AI's integration into daily life. This Claim is backed by Evidence, such as Google's AlphaGo defeating the Go world champion in 2015, a feat previously considered impossible due to the game's complexity. Reasoning explicitly ties this progress to advancements in computational hardware (GPUs), extensive datasets (i.e., Big Data) availability, and breakthroughs in neural network algorithms, particularly deep learning.

Throughout the narrative, the authors examine Implications. Early symbolic "AI Spring" established foundational theoretical principles, yet fell short in solving real-world

complexities. The “AI Winter” emphasized caution and realistic expectations. Expert systems of the “AI Fall” demonstrated specialized competence but lacked flexibility. In “AI Summer,” contemporary deep learning techniques have achieved unprecedented accuracy but introduced new challenges like algorithmic opacity and ethical dilemmas.

The narrative's Context continuously grounds AI developments in historical and technological realities. Examples include Cold War-era research funding shaping early ambitions and the exponential growth in computational power and data availability in recent decades, driving contemporary AI's capabilities.

As this example shows, the CERIC framework provides an efficient means of analyzing a comprehensive historical narrative that articulates the significance of each AI era. The framework allows us to examine how robustly the authors support their Claims with scholarly Evidence, how logically they link developments, what Implications they explore, and how they situate these arguments within a broader historical and societal Context.

SS 12.5 Methodological Evolution

A common narrative strategy, particularly in technical fields of instrumentation or computation, focuses on the evolution of methods. Rather than tracing chronological events, this approach emphasizes how advancements in research approaches—such as new instruments, innovative techniques, or sophisticated analytical strategies—have driven scientific progress. A methodological evolution narrative illustrates how improvements in research practices have expanded our understanding, corrected earlier misconceptions, or opened entirely new investigative pathways.

Let's walk through a hypothetical example of the development of climate change models. The authors might begin by identifying a Claim about the limitations of earlier methodologies. They then provide Evidence drawn from studies conducted during that period to illustrate these constraints—for example, most climate studies in the 1970s relied on simple linear models. Through Reasoning, they explain the shortcomings of these methods—for instance, highlighting how linear models failed to capture complex climate feedback systems, leading to inconsistent or incomplete conclusions. The Implications then clarify that specific critical questions remained unanswered due to methodological limitations.

The narrative would then shift to introduce a new generation of methods, making a fresh Claim such as, "*By the 1990s, computational advances enabled researchers to adopt general circulation models (GCMs).*" This claim would be supported by clear Evidence from landmark studies employing GCMs. The subsequent Reasoning will explain how these advanced models addressed previous methodological drawbacks by incorporating greater detail, more variables, and improved predictive accuracy. The resulting Implications might include significant discoveries that earlier methods could not reveal. Finally, the Context element

situates these advances, noting, for example, how the development of supercomputers facilitated this methodological leap.

Ultimately, a literature review structured around methodological evolution and clarified with the CERIC framework becomes a compelling narrative, demonstrating how each innovation builds logically upon earlier techniques, overcoming limitations and driving the field forward.

SS 12.5.1 Example: Medicine

Consider a review of medical imaging techniques for diagnosing cancer by Fass (2008). The narrative centers on seven major areas of imaging technologies: image contrast; X-ray-based systems, including CT (computed tomography); magnetic resonance imaging (MRI) systems; ultrasound; non-ionizing electromagnetic imaging; nuclear medicine; and positron emission tomography (PET) with CT. Within each area, the review synthesizes the advances and capabilities.

Section 1: Combining Diverse Methods

The narrative begins with a Claim that integrating diverse imaging methods significantly enhances cancer diagnostics. Evidence shows that combining MRI, CT, ultrasound, and nuclear medicine provide complementary anatomical, metabolic, and functional data. The underlying Reasoning is that such integrative imaging approaches yield richer diagnostic insights, which are crucial for accurately detecting cancer and monitoring therapeutic responses. The clear Implications are improved early detection and treatment outcomes. The Context describes rapid advances in bioinformatics and multimodal imaging technologies, facilitating this integrated diagnostic capability.

Section 2: Image Contrast

This section Claims that improving image contrast is essential for precise cancer diagnosis. The review provides Evidence by detailing how intrinsic methods (radiation absorption, magnetic relaxation) and extrinsic contrast agents significantly enhance image clarity. The Reasoning is straightforward: better contrast enables clearer differentiation between tumors and healthy tissues. The major Implications are earlier, more accurate tumor detection and effective treatment monitoring. The technological Context includes advancements in nanotechnology and molecular imaging, which drive innovation in contrast mechanisms.

Section 3: X-Ray-Based Systems Including CT

This section Claims that digital enhancements in X-ray-based imaging, including CT, dramatically improved cancer detection capabilities. Evidence is drawn from advances such as digital radiography, computed radiography, and tomosynthesis, which enhance lesion visualization and differentiation. The Reasoning is that digital techniques overcome the limitations of traditional X-rays, providing superior anatomical detail. The practical Implications include earlier detection, reduced diagnostic uncertainty, and fewer

unnecessary follow-up procedures. The Context is the shift from analog to digital imaging technologies, coupled with the evolution of imaging storage systems.

Section 4: Magnetic Resonance Imaging Systems (MRI)

Fass (2008) makes the Claim that MRI significantly advanced cancer detection and treatment monitoring. The Evidence includes the development of specialized MRI techniques such as perfusion imaging, diffusion-weighted imaging, and spectroscopy. The Reasoning behind MRI's effectiveness lies in its superior soft-tissue contrast and ability to visualize physiological and metabolic processes without ionizing radiation. Implications include enhanced accuracy in diagnosis and targeted therapy, particularly in high-risk patients. The enabling Context involves stronger magnets (i.e., 3 Tesla systems), specialized coils, and significant computational improvements.

Section 5: Ultrasound

The key Claim from this section is ultrasound's critical role in cancer diagnostics due to its non-invasive, real-time imaging capability. The provided Evidence highlights advancements such as elastography, targeted contrast agents, and miniaturized devices. The Reasoning is that these innovations significantly improve ultrasound's diagnostic accuracy by clearly differentiating tumors from healthy tissue. The resulting Implications include earlier tumor detection and reduced reliance on invasive biopsies. The technological Context includes advances in microbubble contrast agents and miniaturized, portable ultrasound equipment.

Section 6: Non-ionizing Electromagnetic Imaging

This section Claims that non-ionizing electromagnetic imaging techniques (e.g., photoacoustic and thermoacoustic imaging) effectively differentiate tumor tissue. Evidence demonstrates that these methods leverage differences in optical and thermal properties between healthy and malignant tissues. The underlying Reasoning is that such methods provide high contrast without the risk of radiation exposure. The significant Implications are safer and earlier cancer detection, especially for superficial or smaller lesions. The developmental Context involves the integration of optical, ultrasound, and radio-frequency technologies.

Section 7: Nuclear Medicine

The primary Claim in this section is that nuclear medicine substantially enhances cancer diagnostics by providing metabolic and molecular imaging. Evidence involves successfully applying radiotracers and Single-Photon Emission Computed Tomography (SPECT) imaging to detect tumors at an early, metabolic stage. The Reasoning is that nuclear imaging uniquely identifies biochemical and physiological abnormalities at a molecular level. Implications include increased diagnostic precision and tailored treatment strategies. The evolving Context encompasses advances in radiotracer chemistry and multimodal imaging techniques that expand clinical capabilities.

Section 8: PET/CT

This section presents the pivotal Claim that PET/CT imaging revolutionized cancer diagnostics by integrating detailed anatomical and metabolic insights. Evidence includes PET's effectiveness in identifying metabolically active cancer regions combined with CT's precise anatomical visualization. The central Reasoning is that combining metabolic and structural data significantly improves diagnostic accuracy and staging precision. The major Implications are earlier interventions, more accurate biopsies, and personalized treatment planning. The enabling Context involves advances in hybrid imaging technology, computational power, and radiopharmaceuticals, collectively propelling PET/CT's diagnostic value.

This example shows how using CERIC explicitly throughout a methodological review can illustrate the ways that each imaging advancement has built logically upon previous methods, driving progress in cancer diagnosis and care.

SS 12.5.2 Example 2: Environmental Science

In their literature review, Malhi et al. (2021) examine the impacts of climate change on agriculture using observational and modeling studies. These approaches are thematically structured around methodology, enabling comparative synthesis.

Observational Studies Theme

A key Claim made by Malhi et al. (2021) is that “*changes in climatic events such as temperature and rainfall significantly affect the yield of crops*” (p. 6). This is substantiated by Evidence from numerous observational and empirical studies cited in their review, including a striking 34% projected reduction in coffee production in Mexico and a measurable decline in net farm revenues in Cameroon attributed to rising temperatures (pp. 6–7). The Reasoning behind these findings is grounded in the observation that observational data reflect real-world, long-term interactions between shifting climate patterns and crop performance, showing that major staples such as maize, wheat, and rice are already experiencing stress under current temperature and precipitation variability. The Implications of this trend are significant: these studies underscore the urgent need for context-specific adaptation strategies, particularly in tropical and arid regions where farmers have limited resources to buffer against climatic shocks. However, the Context of observational studies must be taken into account. They are correlational by nature and often influenced by multiple overlapping variables such as changes in agricultural technology, market conditions, or policy interventions. As such, they provide a valuable but incomplete picture of climate-crop interactions, reinforcing the need for modeling studies that can test specific variables under controlled conditions.

Modeling and Experimental Studies Theme

In their discussion of simulation research, Malhi et al. (2021) offer the Claim that “*simulation models project that increased CO₂ levels may initially improve crop yields, but these gains are*

offset by thermal stress and water shortages” (p. 8). This is supported by Evidence from CERES (Crop Environment Resource Synthesis) simulations in northwest India, which showed yield increases of 28% for wheat and 15% for rice at doubled atmospheric CO₂ levels—benefits substantially diminished when elevated temperatures and water scarcity were included in the scenario (p. 8). The Reasoning behind these results lies in the modeling's ability to isolate and manipulate individual climatic variables, enabling researchers to assess associations and causal relationships between environmental stressors and yield outcomes. These controlled simulations reveal how interacting stressors—particularly heat and water deficits—undermine the otherwise positive effects of CO₂ fertilization.

The Implications are particularly salient for policymakers and agricultural planners: while rising CO₂ might seem like a silver lining for plant growth, effective adaptation will require robust interventions such as irrigation infrastructure, crop genetic improvement, and stress-resilient farming systems. For Context, although crop models like CERES abstract from the full complexity of real farming systems—often omitting socioeconomic, pest, or soil feedbacks—they remain indispensable tools for forecasting and preparing for climate futures, especially in regions projected to experience more intense and frequent extremes.

Comparative Discussion

The review highlights thematic convergence by juxtaposing observational and modeling methodologies. Both show that climate change is reducing agricultural productivity. Observational studies ground the problem in reality while modeling studies extend its understanding into plausible futures. This methodological synthesis, analyzed through the CERIC framework, offers readers a nuanced view of how research design can shape knowledge and recommendations.

12.6 Theoretical Perspectives

Another helpful way to organize a literature review is by using theoretical perspectives or conceptual frameworks. In many fields, researchers examine the same issue through different theoretical “lenses,” each with its own assumptions and ways of explaining it. As Webster and Watson (2002) put it, a strong literature review is conceptual, meaning it groups ideas by theory rather than simply listing studies by author or year. Boote and Beile (2005) make a similar point: the best reviews are grounded in theory and organized around big ideas.

We can see this across many disciplines. In psychology, for example, cognitive, behavioral, and sociocultural theories each offer different explanations for the same behavior. In economics, Keynesian and neoclassical models interpret markets in very different ways. These examples show how theory shapes what researchers pay attention to and how they make sense of evidence.

When a field has clear theoretical “camps,” organizing a review around those perspectives can make your writing clearer and the synthesis stronger. It helps readers see how each framework contributes to understanding the problem—and where the perspectives agree, disagree, or leave questions unanswered. In short, grouping sources by theory makes it easier to show how different viewpoints shape what we know.

Using CERIC within each subsection deepens the analysis, by mapping the elements to key review points, as follows:

- **Claim:** What the theory asserts
- **Evidence:** Studies that support the theory
- **Reasoning:** How the theory explains the findings
- **Implications:** What the theory helps predict or overlook
- **Context:** The theory’s origins, scope, and critiques

To illustrate how this works, Table SS 12.1 compares two behavioral theories as understood through the CERIC framework.

Table SS 12.1 Comparison of CERIC Elements in Two Behavioral Theories

CERIC Framework	Social Identity Theory (Tajfel & Turner, 1979)	Rational Choice Theory (Becker, 1976)
Claim	Individuals behave in line with perceived group membership.	Individuals act based on cost-benefit calculations.
Evidence	Studies in teams, organizations, and online communities show conformity to in-group norms.	Research on norm-breaking, whistleblowing, and strategic behavior.
Reasoning	Salient identities guide behavior to preserve social belonging.	Utility maximization often overrides group conformity when incentives are high, depending on the strength of identity (which itself can be part of utility) and social preferences (e.g., fairness, reciprocity)

Implications	Explains cohesion but may falter when identity isn't salient.	Useful for explaining self-interested behavior; limited in capturing social/emotional drivers.
Context	Rooted in social psychology, refined through decades of empirical testing.	Originated in economics; extended into political science and organizational theory.

When readers encounter a literature review structured by a theoretical perspective and synthesized using CERIC, they are invited into a scholarly dialogue between frameworks. This approach clarifies how different theories shape research design and interpretation and models the critical thinking necessary to navigate, compare, and evaluate diverse strands of academic inquiry.

Let's take a deeper dive into a specific literature review example.

SS 12.6.1 Example: Ecology

Austin's (2007) literature review examines species distribution models and ecological theories. The review is thematically structured around theoretical and conceptual frameworks, enabling comparative synthesis by major theme.

Theme 1: Species Response Curves and Niche Theory

A main Claim in Austin (2007) is that “*species’ responses to environmental gradients are typically unimodal and often skewed*” (p. 1). The Evidence for this is grounded in both empirical studies and theoretical expectations from niche theory. Austin (2007) notes that many species distribution models (SDMs) employ polynomial functions that fail to capture realistic ecological shapes—especially skewed or truncated responses. The Reasoning here is based on the idea that species’ physiological tolerances and interactions with biotic and abiotic factors give rise to nonlinear and asymmetric distributional responses. Austin critiques current practice by stating that “*explicit theory needs to be incorporated into species response models used in conservation*” and urges testing for unimodal skewed responses as a routine procedure (p. 1). The Implications are significant: default model assumptions may lead to ecologically implausible predictions, particularly at the tails of species’ environmental ranges. The Context is that niche theory has guided ecological modeling for decades, but is often applied in oversimplified forms in statistical models without explicit theoretical justification.

Theme 2: Scale and Data Resolution

Austin (2007) also claims that the resolution and extent of data collection “*can severely restrict the purposes for which the data can be used*” and influence the shapes of species responses that can be detected (p. 4). The Evidence includes examples from studies that model the distributions of New Zealand tree species, in which high-resolution plot data (0.4 ha) enabled assessment of competition and climate effects. In contrast, coarser datasets (e.g., 50 km grid cells) could only detect broad climate signals. The Reasoning is that if the sampling range fails to encompass the full environmental gradient, observed species responses will be truncated, preventing accurate inference of the true ecological niche. The Implications for model interpretation are critical: data resolution determines whether biotic interactions, microclimate effects, or spatial dependencies can be captured, thereby directly shaping conservation and forecasting efforts. The Context reflects long-standing challenges in spatial ecology, where data limitations and scale mismatches often compromise ecological realism.

Theme 3: Environmental Predictors and Conceptual Frameworks

Another Claim in the article is that predictor variables should be chosen based on their ecological and biophysical relevance, following a framework that distinguishes between indirect, direct, and resource variables. Evidence supporting this includes examples showing that using slope and aspect (indirect predictors) instead of radiation (a direct predictor) can lead to different, sometimes irrational response curves. The Reasoning follows from recognizing that predictors with known causal relationships to species performance—such as soil moisture or nutrient availability—support stronger mechanistic interpretations of response curves. The Implications are that model selection and predictor inclusion should be theory-driven rather than based solely on availability or statistical fit. Context is provided by earlier frameworks such as Austin and Smith (1989), who called for more biologically informed modeling approaches—an appeal that remains widely unheeded.

Theme 4: Quantile Regression and Limiting Factors

Austin (2007) presents quantile regression as a promising approach consistent with Liebig’s Law of the Minimum, which states that a single limiting factor can constrain species performance even when other factors are sufficient. The Claim here is that traditional regression methods, which estimate mean responses, obscure these limiting effects and lead to misleading inferences. Evidence includes case studies showing how upper quantile regressions capture physiological limits more accurately than means, particularly in observational datasets influenced by multiple stressors. The Reasoning is that quantile regression can estimate species’ performance envelopes and detect ecological thresholds, where limiting factors dominate. Implications include improved habitat modeling, clearer identification of critical constraints, and enhanced forecasting of species responses to environmental change. Context is provided by the author’s broader critique of statistical practices in ecology: quantile regression is part of a new class of methods that better aligns with ecological processes and should be more widely adopted in SDM practice.

Overall, the CERIC method strengthens the review's structure and equips readers to evaluate how different theoretical frameworks generate, justify, and challenge knowledge. These skills are crucial when engaging with complex theoretical literature.

SS 12.7 Establishing a Narrative Arc

Narrative arcs are common ways of organizing literature reviews, where advances are analyzed around major ideas.

SS 12.7.1 Example: A Debate in Medical Science—Sodium and Hypertension

Let's say you are reviewing a long-standing debate in medical literature: Does dietary sodium cause hypertension?

Position A: Sodium is a Primary Cause of Hypertension

- **Claim:** High sodium intake significantly contributes to elevated blood pressure.
- **Evidence:** Multiple clinical trials demonstrate that reducing sodium intake lowers blood pressure, particularly in older adults. Landmark studies like the Dietary Approaches to Stop Hypertension (DASH) trial support this association (Conlin et al., 2000).
- **Reasoning:** Researchers argue that sodium increases blood volume via fluid retention, which in turn raises pressure within the vascular system. The strength of this Reasoning is often supported by controlled conditions and large sample sizes.
- **Implications:** If true, this work supports the widespread adoption of dietary guidelines recommending reduced sodium intake for cardiovascular health.
- **Context:** Many of these studies focus on older or at-risk populations. While this interpretation has shaped mainstream medical advice for decades, it is increasingly being re-examined in light of new evidence.

Position B: Sodium's Impact is Minimal or Highly Individualized

- **Claim:** For most people, sodium has little or no meaningful impact on blood pressure.
- **Evidence:** Large-scale epidemiological studies have shown no consistent association between sodium intake and hypertension across diverse populations (Intersalt Cooperative Research Group, 1988). Some genetic research suggests only certain "salt-sensitive" individuals are affected.
- **Reasoning:** Supporters of this position argue that earlier studies were confounded by other dietary or lifestyle variables (e.g., processed food consumption correlates with multiple risk factors). They also critique the duration of clinical trials, suggesting short-term interventions might not reflect long-term physiological regulation.
- **Implications:** If this position holds, public health recommendations may need to shift from universal sodium reduction to more personalized dietary guidelines.
- **Context:** This view is more recent and challenges long-standing nutritional orthodoxy. It has sparked significant debate and re-analysis of earlier data.

By applying CERIC to each side of a controversy, the review becomes a structured conversation. Each Claim can be quickly examined, each line of Evidence weighed, and each Reasoning pathway mapped. Implications are clearly drawn, and Context keeps the reader anchored in the field's evolving landscape. When done well, a controversy-centered review models the process of scientific disagreement and shows how knowledge develops through tension, refinement, and revision.

Let's take a deeper dive into a specific literature review example.

SS 12.7.2 Example: Astronomy

Literature on the search for extraterrestrial life often highlights a scientific controversy: the interpretation of possible biosignatures in Martian meteorites. Gibson et al. (2001) exemplified this debate by reassessing their original claim from McKay et al. (1996) in light of new data.

Position A

- **Claim:** Morphological, mineralogical, and chemical features in Martian meteorites ALH84001, Nakhla, and Shergotty are consistent with ancient microbial life on Mars.
- **Evidence:** Gibson et al. (2001) cite multiple converging lines of evidence. These include low-temperature carbonate globules in ALH84001, magnetite crystals matching those formed by terrestrial magnetotactic bacteria, indigenous reduced carbon compounds, and structures morphologically similar to bacterial cells, colonies, and biofilms found in all three meteorites. Importantly, some of these features occur in sealed interior regions, minimizing the likelihood of contamination. The detection of polycyclic aromatic hydrocarbons (PAHs) and isotopic signatures consistent with biological carbon fractionation further support the claim.
- **Reasoning:** The authors argue that although individual features (e.g., magnetite or organics) may have abiotic explanations, their spatial co-location within the same microenvironments supports a biogenic interpretation. The combination of bacterial-like morphology, biofilm-like coatings, biogenic magnetite, and reduced carbon compounds indicates microbial activity. The authors reason that this convergence of indicators is most parsimoniously explained by biogenic processes, especially since comparable features have not been found in non-Martian meteorites or lunar samples examined with similar scrutiny.
- **Implications:** If correct, this interpretation would represent the first evidence of life beyond Earth, suggesting that Mars may have hosted microbial ecosystems during its warmer, wetter past and perhaps still harbors life today in subsurface niches. This would reshape astrobiology, planetary protection protocols, and mission design for Mars exploration.
- **Context:** The initial 1996 announcement generated global excitement and intense scientific scrutiny. Gibson et al. (2001) acknowledge extensive challenges to their

interpretation, ranging from alternative abiotic explanations to concerns over contamination. However, their updated synthesis incorporates findings from newly studied meteorites and improved analytical techniques, illustrating how scientific understanding evolves through iterative investigation.

Position B

- **Claim:** The features in ALH84001 and related meteorites are the result of abiotic mineral and chemical processes, not evidence of life.
- **Evidence:** Critics, including Bradley et al. (1996, 1997), Scott (1999), and others, re-examined the meteorites using high-resolution microscopy and geochemical models. They argue that the magnetites could be shock products, that the PAHs might result from Antarctic contamination, and that the morphological features are below the lower size limit for viable life or are consistent with known mineral artifacts. Some sulfur and carbon isotope studies also failed to show definitive biogenic fractionation.
- **Reasoning:** Skeptics maintain that each line of Evidence is ambiguous on its own and that extraordinary claims demand extraordinary evidence. They emphasize that alternative abiotic processes (such as vapor phase crystallization, carbonate precipitation, or terrestrial contamination) can produce similar structures. For example, some features may result from terrestrial weathering or sample preparation artifacts.
- **Implications:** The more conservative conclusion is that no unambiguous evidence of Martian life exists yet. Accepting a biogenic interpretation prematurely could lead to flawed assumptions about Mars' habitability and misguide mission priorities. Instead, future exploration should focus on in situ sampling and rigorous biosignature criteria.
- **Context:** This controversy highlights the scientific method in action—hypotheses, tests, refutations, and ongoing refinement. It also underscores the necessity for stringent contamination controls and robust criteria for detecting life. The debate catalyzed NASA's Astrobiology Program and has influenced sample-return mission planning for decades to come.

Often, after presenting multiple sides of a controversy, a review will synthesize the current state of a debate. This review may involve noting where a consensus is emerging, highlighting lingering tensions, or even identifying a third perspective reconciling opposing views (UCLA Undergraduate Writing Center, 2016). In the case of Martian meteorites, the review might conclude that while current evidence for microbial life is likely abiotic, the search for biosignatures continues—now expanded through new missions using rovers and drills to detect organic compounds. Such a resolution doesn't erase the controversy but places it within a broader scientific trajectory.

SS 12.8 Software Tools for Efficient and Ethical Review Construction

Modern researchers have a wealth of software tools to streamline the literature review process. These tools can save time, improve organization, and even help uphold ethical

standards (for example, by keeping a careful track of sources to avoid inadvertent plagiarism). Here, we provide an overview of key tool categories and their applications at the time of publication of this book, with emphasis on the natural and applied sciences.

SS 12.8.1 Reference Management Software

Applications like *Zotero*, *EndNote*, and *Mendeley* are highly valuable for collecting and organizing citations. They allow you to import references directly from databases or the web, store PDFs, and annotate them. Critically, they integrate with word processors to cite while you write, and automatically build bibliographies in your chosen citation style. Using a reference manager helps ensure you cite sources accurately and consistently, which is both efficient and ethical. For instance, instead of manually tracking dozens of references, a tool like *Zotero* can manage them and insert properly formatted citations, reducing the risk of mis-citing. After gathering papers you plan to include in the review, store them in your preferred reference manager (*EndNote*, *Zotero*, *Mendeley*, etc.). These tools also support tagging or categorizing literature, which can map to themes in your review. In STEM fields, where one might juggle hundreds of papers, reference managers are essential for efficiency.

SS 12.8.2 PDF Annotation and Notetaking Tools

As you read your sources, it's helpful to annotate them and take organized notes. PDF readers like *Adobe Acrobat* or *PDF-XChange* allow highlighting and commenting directly on PDFs. More specialized tools (e.g., *LiquidText*, *MarginNote*, or *Hypothes.is*) enable layering notes and linking across documents. Note-taking apps such as *Evernote*, *OneNote*, or *Notion* let you compile summaries and quotes in a searchable format. The key is to maintain an annotated bibliography or notes database as you go – this fosters ethical writing because you're less likely to accidentally plagiarize if you have clearly marked notes of others' ideas. Some students use cloud-based notebooks or spreadsheets for this. For example, one structured approach is using a spreadsheet or table to record for each source: its main Claim, methodology, key findings, and your thoughts. The CERIC framework fits well with such tools, prompting you to record specific elements for each paper. By systematically annotating, you can later synthesize the literature more easily, ensuring your review is grounded in the actual content of sources rather than copy-pasted text.

SS 12.8.3 Mind Mapping and Visualization Tools

To organize ideas and see connections between studies, many researchers turn to mind maps or visual tools. Software such as *MindMeister*, *XMind*, or *CmapTools* allows you to create concept maps of themes, theories, and findings. In STEM, visualization can be especially useful to group studies by method or to map out how one study leads to another. Another class of tools uses citation data to generate visual literature maps. For example, *Connected Papers* and *Research Rabbit* create graphs of related papers based on citation networks. These can help you discover relevant work you might have missed and understand the

intellectual lineage of a topic (which papers are seminal, which clusters of papers form subtopics, etc.). Using such tools ethically means still reading and critically evaluating the sources – but they provide a big-picture view that can guide how you structure your review. A mind map of your own project’s literature, for instance, can ensure you cover all major areas and see how they link, leading to a more coherent and comprehensive write-up.

SS 12.8.4 Grammar and Style Checkers

While not specific to literature reviews, general writing aids like *Grammarly*, *Hemingway Editor*, or *Microsoft Word’s* editor can improve the readability of your review. They catch grammar mistakes, overly complex sentences, and overuse of passive voice. This contributes to an accessible writing style. Some tools also have plagiarism detection features (e.g., *Grammarly’s* premium version) which can double-check that your phrasing isn’t too close to source material – an added ethical safeguard if used properly. In the natural sciences, where authors may not be writing in their first language or are aiming for publication, these tools help polish the text. Just use them judiciously: they are best at suggesting improvements, but the writer must decide what to keep accurate. Also, ensure that any AI-based editing tool used on sensitive drafts complies with data privacy or journal policies (some journals ask that you disclose if a grammar-correcting AI was used in preparation of a manuscript, though this is still an evolving area).

SS 12.8.5 Structured Frameworks (e.g., CERIC) Plus Tool Integration

A structured approach like CERIC can greatly enhance both efficiency and rigor in a literature review. The CERIC framework prompts the reviewer to extract and record the five elements of Claim, Evidence, Reasoning, Implications, and Context from each paper, ensuring that you critically engage with sources rather than just summarizing. Tools like *Google Forms* or *Notion* can be customized to capture CERIC elements for each article (Bjorn et al., 2022). For example, a student might use a *Google Form* for each paper they read, filling fields for Claim, Evidence, etc., which then feeds into a spreadsheet for analysis. This method, combined with social collaboration tools such as *hypothes.is* has been used in graduate courses to build a shared annotated database of literature (Bjorn, 2023).

The integration of a structured framework with software means you leverage the best of both: human insight for analyzing each source and computer assistance for organizing and retrieving that analysis. These combined approaches also help maintain a high-level view of the literature and support ethical writing by keeping your interpretations tied to specific recorded evidence from sources. These tools further allow researchers to manage large volumes of literature efficiently, saving time and reinforcing ethical practices like note-taking and accurate citation. However, it’s important to remember that tools are aids, not substitutes for critical thinking. They support the process, but ultimately, the researcher’s analytical skills ensure the literature review is meaningful and ethically composed.

SS 12.9 Generative AI and the Literature Review Process

Generative artificial intelligence (GenAI) is reshaping how academic writers approach the literature review process, particularly in the sciences. Whether brainstorming questions, finding relevant sources, or summarizing findings, GenAI tools such as *ChatGPT*, *Elicit*, and *Scite* offer real opportunities to improve writing fluency and research efficiency. When used wisely, these tools act as assistants—enhancing discovery, structure, and clarity—but must be guided by the user’s critical thinking and scholarly integrity. This section introduces some widely used GenAI tools at the time of publication of this book, explains how to apply the CERIC method to sharpen their use, and offers guidance on avoiding common pitfalls and ethical violations. See Chapter 14 and its associated supplement for more information on how GenAI tools can be used effectively and ethically to engage with the literature.

SS 12.9.1 Key Generative AI Tools for Literature Reviews

A growing ecosystem of AI tools—both academic-focused and low-cost commercial—is available to support literature reviews. While most tools have strong coverage in STEM fields, interdisciplinary applications are growing. Researchers should supplement the use of GenAI tools with discipline-specific databases and librarian expertise. Let’s review some of the current AI tools most applicable to higher education, along with CERIC integration tips and boundaries for ethical use and avoiding overreliance.

Figure SS 12.1: Summary of Generative AI Tools for Literature Reviews

Summary of GenAI Tools for Literature Reviews

- **ChatGPT and LLM-based Assistants:** General-purpose large language model chatbots that can answer questions, generate summaries, or compose text. These are often used for brainstorming and drafting.
- **Consensus:** An AI-powered academic search engine that answers specific questions by retrieving relevant literature and indicating consensus.
- **Elicit:** An AI research assistant focused on literature search and synthesis. It finds relevant papers, summarizes key takeaways, and extracts pertinent information from abstracts.
- **Litmaps:** Uses citation networks to recommend relevant literature and creates visual maps of how papers cluster over time.
- **Scite Assistant:** Combines natural language answers with citation analysis, using smart citations to classify how a paper is cited by later studies.
- **Semantic Scholar AI, Connected Papers, and Research Rabbit:** Visualizes networks of papers, identifies influential works and clusters of research, and suggests related works interactively.
- **SciSpace, Perplexity AI, and EvidenceHunt:** Provides AI assistants that can answer questions from uploaded PDFs, cites sources in responses, and specializes in domain-specific literature.

SS 12.9.1.1 ChatGPT (and Similar Large Language Models)

ChatGPT is a widely used generative AI tool that produces text in response to user prompts. In the context of literature reviews, it can support early-stage tasks such as brainstorming research questions, exploring background concepts, or drafting provisional summaries. For instance, prompting ChatGPT with “*What are key themes in renewable energy storage literature?*” may yield a helpful outline to refine further. It can also assist with sentence revision and flow, offering clearer phrasing for rough drafts. ChatGPT’s strength lies in its flexible, conversational interface, which enables iterative refinement.

However, ChatGPT does not have direct access to academic databases, and its responses are based on training data and possibly old research literature. As a result, it may omit recent developments or generate inaccurate or fabricated content. Researchers should treat their output as provisional scaffolding to be verified against primary sources—never as a substitute for critical reading or evidence-based synthesis.

CERIC Integration Tip: A useful approach for ChatGPT is to test or generate structured prompts that target specific CERIC elements of your literature. For example: “*What is a possible reasoning link between elevated CO₂ and reduced stomatal conductance in plants?*” This helps you articulate your own Reasoning or Implications more clearly. In a review of plant physiology, this might help you connect experimental findings to a broader Context in climate science.

Ethical Use Boundary: Using ChatGPT to generate ideas, draft outlines, or refine your own wording is considered ethical in many (but not all) areas of natural science research; however, it is neither ethical nor prudent to present its unverified output as your own analysis or factual summary. For example, you may ask ChatGPT, “*What are some possible subtopics in the literature on microplastics?*” and use the response to spark further search and reading. However, if ChatGPT gives you a Claim like “*A 2021 study found that microplastics reduce phytoplankton photosynthesis,*” you should verify that this study exists in the literature and the Claim is stated correctly before including it in your review.

Overreliance Boundary: Do not let ChatGPT do the intellectual heavy lifting of evaluating or synthesizing your sources. For example, you can ask it to rephrase a rough draft paragraph for clarity, but avoid asking it to “*summarize these five papers and write a literature review section.*” Doing the latter skips the deep engagement with content necessary to form your own understanding and judgment.

SS 12.9.1.2 Elicit

Elicit (Ought.org) is an AI research assistant designed to support the early stages of the literature review process, particularly in the natural and applied sciences. Unlike traditional search engines, *Elicit* uses semantic search to identify relevant papers based on a user’s research question rather than relying solely on keyword matching. For example, asking “*What*

are the effects of microplastics on marine life?” may return papers that use different terminology but address the core question, along with brief, synthesized summaries. Importantly, *Elicit* provides references for its answers, often with direct links to the papers, which is crucial for verification. This functionality makes *Elicit* especially useful during the literature discovery and early synthesis phases. It can surface interdisciplinary sources that may be overlooked and provide reference-linked outputs for further reading. While its summaries offer a starting point, researchers must still read and critically assess each paper. At best, it can accelerate the identification of which papers to read.

CERIC Integration Tip: As you review search results, tag or group papers by Claim (e.g., “CO₂ boosts yield”) and note what kind of Evidence each study uses (e.g., greenhouse experiments vs. field trials). Then extract the Reasoning and Implications from the authors’ discussions. For example, in agricultural climate impact studies, comparing Evidence across CO₂-focused vs. drought-focused papers can help define subthemes for synthesis.

Ethical Use Boundary: In research, it is generally appropriate to use *Elicit* to search for and organize literature, but do not cite its summaries directly without reading the full articles yourself. For example, if *Elicit* provides a summary of a paper’s findings on crop yield under drought conditions, treat it as a lead. You must read the article and determine if the summary accurately reflects the authors’ conclusions before integrating it into your synthesis.

Overreliance Boundary: Do not rely on *Elicit* to determine the strength or validity of Evidence and Reasoning—that’s your responsibility. For example, *Elicit* may surface five studies on a given topic, but you still need to assess methodological quality, relevance, and agreement between findings. While very tempting, avoid using *Elicit*’s “key takeaways” as a substitute for critical reading.

SS 12.9.1.3 Scite

Scite is an AI-powered citation analysis platform supporting evidence evaluation and literature review synthesis. It distinguishes between citations that support, contradict, or merely mention a given source, allowing researchers to trace how a study has been received in subsequent literature. This is particularly helpful when identifying areas of consensus or controversy within a research field. Scite’s conversational assistant can also respond to queries like “*What do studies say about X?*” by returning citation-backed summaries drawn from its database. Because its outputs are tied to verifiable citation contexts, Scite offers a more transparent alternative to tools like ChatGPT that can sometimes generate fabricated Claims. In STEM fields, especially, where determining the strength and trajectory of Evidence is essential, *Scite* can enhance critical analysis by helping reviewers anchor their synthesis in the actual scholarly record.

CERIC Integration Tip: Use Scite to track how specific Claims evolve. Start by identifying a key Claim, then review the Reasoning in papers that support or contradict it. For example, if a

review focuses on coral bleaching mechanisms, you can follow which studies reinforce or challenge the Implication that acidification is a primary driver, and document how the Context shifts across citations.

Ethical Use Boundary: Scite’s citation context labels (supporting, contrasting, mentioning) can help you better understand scholarly conversation; however, do not quote or interpret them directly without checking the actual citing paper. For example, if Scite says a paper is frequently “contradicted,” examine those contradictory studies yourself to ensure the disagreement is substantive and not limited to a select set of studies or specific sub-claims.

Overreliance Boundary: Don’t assume Scite’s metrics or labels capture the full complexity of scholarly consensus or debate. For example, a paper may be labeled as “supporting” many times, but those citations could be superficial. Use *Scite* to prompt deeper inquiry, never as a final verdict on the literature’s position.

SS 12.9.1.4 Research Rabbit (and Other Literature Mapping AI Software)

Research Rabbit is a literature-mapping tool that visualizes high-level connections among papers, authors, and research clusters. By inputting a few key papers or keywords, users can explore networks of related work, identify collaboration patterns, and discover recommended sources through machine-learning-driven associations. Though it does not generate text, its visual interface excels at revealing conceptual links that keyword searches might miss. This functionality is particularly valuable in interdisciplinary research, where methods or findings from one field may have applications in another. For example, Research Rabbit might reveal that a hydrology modeling technique has also been adopted in climate forecasting—connections that traditional searches may not surface. Used alongside tools like Elicit or Scite, Research Rabbit expands the scope of literature discovery and supports a more integrative search by illustrating how bodies of research interrelate.

CERIC Integration Tip: Color-code or annotate nodes in the map based on Claim types (e.g., mechanistic vs. correlative) or dominant types of Reasoning. For example: In an SDM (Species Distribution Model) review, group one set of nodes around papers using correlative models (each with different Reasoning) and another around mechanistic models, then compare their Implications for predicting range shifts.

Ethical Use Boundary: Use Research Rabbit to identify clusters of relevant literature and trace citation networks, but do not assume proximity in the graph means conceptual alignment or quality. For example, two papers may appear closely connected on the map but have opposing conclusions. Always read abstracts (and ideally full texts) before including them in your review.

Overreliance Boundary: Don’t let the tool define the scope of your literature review. For example, you might use the literature map to explore new directions, but supplement this

practice with traditional keyword searches and expert consultation. Relying solely on *Research Rabbit* could cause you to miss newer or less cited but relevant studies.

SS 12.9.1.5 Other GenAI-Aided Tools

Beyond the tools covered above, a range of additional genAI applications can support specific stages of the literature review process. Tools like Paper Digest and Scholarcy generate concise outlines of individual articles, while platforms such as IBM Watson Discovery and GPT-based plugins enable custom searches within targeted document sets. Even reference managers like Zotero now incorporate genAI features to suggest related readings, enhancing discovery within curated libraries. General-purpose tools, such as Google's Generative AI in Search or Microsoft Bing AI, can assist with technical queries and occasionally provide citation-linked responses. While these tools may serve more niche roles, they can complement a workflow when used strategically.

CERIC Integration Tip: Use AI summaries to populate a quick matrix where you extract each paper's Claim, type of Evidence (e.g., in vitro, in vivo, simulation), and Implications. Then, refine that matrix by adding your own Reasoning and Context notes after reading the full article. For example, in biomedical literature on mRNA vaccine development, this process helps you sort clinical trials from animal models, clarify assumptions, and identify ethical debates.

Ethical Use Boundary: It is helpful to use summary generators or TLDR ("Too Long, Didn't Read") tools to orient yourself to a paper, but always verify with the full text before citing, paraphrasing, or drawing conclusions. For example, a TLDR might say "*This study finds that X increases Y,*" but the full paper may include significant limitations or caveats. Never cite a finding you've only seen in summary form.

Overreliance Boundary: Do not use these tools to replace your own annotation and note-taking process, which is vital for comprehension and synthesis. For example, use Scholarcy to get a quick overview, then manually annotate the PDF and take your own structured notes (e.g., using the CERIC method) to ensure real understanding.

SS 12.9.2 How GenAI Can Help: Use Cases Across the Literature Review Lifecycle

Generative AI and related tools can be slotted into various stages of conducting a literature review. Let's take a deeper dive into some of the best use cases.

SS 12.9.2.1 Brainstorming and Scoping

At the start of the literature review process, GenAI tools can help explore and refine research questions. ChatGPT, for example, can suggest subtopics, pose possible research questions, or outline themes in a given area (e.g., "*What are emerging themes in renewable energy storage?*"). These outputs can serve as a springboard for idea generation and help define the

conceptual boundaries of a topic. GenAI tools can also generate provisional background overviews—for instance, summarizing the development of a theory or technique. While all factual details must be independently verified, using AI at this stage can help overcome blank-page anxiety and provide a structured entry point into the literature. As always, any factual details from these brainstorming need to be checked and confirmed.

CERIC Integration Tip: As you brainstorm, prompt AI tools to suggest preliminary Claims or areas of Implication for further exploration. Ask: “*What are common explanations for the decline in pollinator populations?*” or “*What Contexts have recent studies used to frame battery innovation?*” These queries help you generate leads that should be checked and could become structured entries in your review.

SS 12.9.2.2 Literature Discovery

Once a research focus has been established, genAI tools can accelerate the process of discovering literature. Elicit, for example, enables question-based searches that yield studies relevant to specific prompts. Tools like Research Rabbit and Connected Papers visualize citation networks and related clusters, helping to uncover sources that may not surface through standard keyword searches. Even ChatGPT can suggest key researchers or landmark studies, which can be validated through academic databases. Some researchers use AI-assisted alerts to stay current on new publications in their field, making this phase more efficient and responsive. These tools are most effective when paired with traditional search strategies, like Google Scholar and library databases. As always, each source requires critical evaluation to ensure its relevance to the review.

CERIC Integration Tip: As you gather sources, classify them by potential Claim and type of Evidence (e.g., observational, modeling, or experimental). Ask genAI tools to help you extract core Implications or Context—for example, “*What do these papers suggest about future directions in ocean acidification research?*” This allows you to build a literature map organized by CERIC dimensions from the outset.

SS 12.9.2.3 Summarizing and Thematic Analysis

As you move into the synthesis stage of the literature review, generative AI tools can assist with organizing, summarizing, and structuring information. Tools like Scholarcy and ChatGPT can quickly condense long papers into key points, which is particularly useful during the initial screening phase. Once you’ve identified relevant studies, you can prompt *ChatGPT* to compare main ideas across papers, highlighting areas of agreement, contradiction, or emerging patterns. For larger collections of sources, AI-enhanced reference managers and clustering tools can sort papers by methodology, topic, or theoretical framework—helping you detect structure in the evidence base. For instance, if you’ve accumulated a series of notes or excerpts, you might ask a generative AI tool to suggest a logical organizational scheme (e.g., “*What could be the major sections of my literature review based on these findings?*”). The tool might return a draft structure, such as background context, methods

used, results by subtopic, and areas of controversy, providing a helpful starting point for refinement. GenAI can also serve as a drafting assistant during early synthesis, offering phrasing suggestions or rewording rough paragraphs to improve clarity and cohesion. These functions can support fluency and consistency in tone. As always, all AI-generated outputs must be verified and edited to ensure accuracy and avoid misrepresentation or mis-citation. When used with careful oversight, AI tools can be a valuable sounding board—helping to connect findings across studies while preserving the integrity and originality of the analysis.

CERIC Integration Tip: During synthesis, prompt AI to help identify recurring Reasoning patterns across studies or to compare the Implications of two competing Claims. For example, ask: “*What is the main reasoning used in studies supporting vs. questioning the hygiene hypothesis in allergy research?*” This helps you construct comparative sections grounded in the CERIC framework.

SS 12.9.2.4 Drafting and Editing the Review

In the writing phase, genAI can help produce a draft text or improve an existing one. Some people experiment with having ChatGPT draft a portion of the literature review—for example, “*Write a paragraph summarizing recent advances in battery technology based on these notes...*”. If attempted, this should be done with extreme caution. AI-written text can sound fluent but may include subtle inaccuracies or even hallucinated citations if asked to cite (more on that risk in the next section). A safer use is to have genAI help with micro-tasks, such as suggesting a transition sentence, a simpler way to state a complex idea, or checking grammar and tone consistency. AI can also assist with formatting citations if integrated properly, although reference managers are typically more reliable for this work. After drafting, one might use an AI-based tool to proofread and flag unclear sentences. In essence, genAI can act like a junior writing assistant by providing suggestions and temporary text that the human author then fact-checks, edits, and integrates as appropriate. This collaboration can speed up the writing process, but the author remains in full control of the final content to maintain academic integrity and accuracy.

CERIC Integration Tip: Use CERIC as a quality control checklist during revision. As you refine paragraphs, ask: “Is the Claim clearly stated? Is the Evidence specific and cited? Does the Reasoning connect them logically?” You can also ask AI to highlight missing Implications or prompt you to clarify Context—for instance, “*How have interpretations of this finding changed over time?*”

SS 12.9.3 Benefits and Productivity Enhancements

When thoughtfully integrated into the research process, generative AI tools can offer substantial productivity gains for students and scholars conducting literature reviews. These tools do not replace the intellectual work of synthesis. Still, they can accelerate routine tasks, expand access to diverse sources, and support skill development in critical reading and

academic writing. For graduate students in particular, AI-enhanced workflows can relieve time pressure and cognitive overload, enabling deeper engagement with ideas.

This section highlights five key benefits—efficiency, idea generation, literature expansion, consistency, and learning—and shows how each can be strengthened by aligning with the CERIC framework.

SS 12.9.3.1 Efficiency and Speed

AI can drastically cut down the time spent on certain tasks. Searching for relevant literature that might take days can be accelerated with intelligent query-answering tools like *Elicit*. Summarizing a 20-page article might take a human several hours, but an AI could produce a rough summary in minutes or seconds. This efficiency enhancement lets researchers cover more ground quickly, which is particularly beneficial in fast-moving fields or for graduate students facing tight deadlines.

As an example, an AI-generated summary can quickly identify the Claim and Evidence in an article, allowing more time to critically engage with Reasoning, Implications, and Context. Speed enhances your ability to process large bodies of work while still applying a structured, critical lens.

SS 12.9.3.2 Idea Generation and Overcoming Roadblocks

AI's ability to generate text and ideas can help overcome writer's block or spark new perspectives. If you are struggling with how to introduce your review, asking ChatGPT for an intro paragraph or an outline can provide a creative jumpstart (to be edited and tailored by you). The AI might highlight connections you hadn't thought of or present a counterpoint that you can then address. In interdisciplinary reviews, where you might not be an expert in all subfields, AI can also help explain unfamiliar concepts, acting as a tutor. AI brainstorming can also help you predict initial Claims from the Evidence, explore alternative Reasoning strategies, or pose questions about missing Implications, offering an accessible way to initiate structured critical thinking across disciplines.

SS 12.9.3.3 Broadening Literature Scope

Tools like Research Rabbit and Elicit can uncover papers outside of your immediate search terms, ensuring a more comprehensive review. This is useful in STEM, where relevant literature might span sub-disciplines. For instance, a materials science innovation might be discussed in chemistry journals and engineering conferences. AI can help bridge those silos by focusing on content relevance. By catching diverse sources, AI contributes to a more robust synthesis that accounts for multiple viewpoints. A wider scope of literature enables you to compare more Claims across disciplines, weigh different types of Evidence, and identify areas where Reasoning or Implications diverge—strengthening both thematic richness and analytical depth.

SS 12.9.3.4 Consistency and Attention to Detail

AI can help maintain consistency in terminology and style throughout the review. For instance, if you describe a concept in one section, you can ask the AI to ensure that a later section uses the same terminology, or to remind you if a term hasn't been defined. AI can also track details like acronyms and ensure that all points you intended to cover are mentioned. For example, asking “*Have I addressed the main arguments of authors A, B, C in this draft?*” may identify if you omitted something you had partially discussed in the draft. This approach can complement your own revision process to catch omissions or inconsistencies. GenAI can also help ensure each paragraph contains a clear Claim, that Evidence is cited consistently, and that the Context for each study is properly introduced—reinforcing logical coherence and completeness across the review.

SS 12.9.4 Limitations and Common Pitfalls

Despite its advantages, genAI has significant limitations and risks that must be acknowledged and dealt with when used for literature research. Figure SS 12.2 summarizes the critical limitations to keep in mind when using these tools.

Figure SS 12.2. Critical Limitations of GenAI for Literature Reviews

GenAI's Critical Limitations
<ul style="list-style-type: none">• Academic integrity and ethical violations: Using genAI in academic work raises questions of transparency and fairness, including the risk of plagiarism.• Access issues: Tools often lack access to paywalled or proprietary literature, excluding potentially large swaths of relevant source information.• Bias in output: Bias in the training data available for genAI tools may result in skewed perspectives.• Hallucinated references: AI may fabricate citations and information. All references and summarized findings must be independently verified.• Inconsistency: GenAI produces varying or inaccurate outputs across sessions and platforms, making it essential to go back to original sources.• Outdated information: Some tools have limited knowledge of recent literature depending on the timing of training. Always supplement AI searches with traditional literature searches to find recent results.• Overreliance: Excessive dependence on genAI can weaken core academic skills, such as critical thinking and deep understanding of a field.

SS 12.9.4.1 Hallucinated Citations and Facts

A well-documented issue is that tools like ChatGPT can fabricate realistic-sounding references or facts that are not real (Dolan, 2024), with false citation rates in some tests ranging from 6% to 60% (Elmslie, 2023; MacDonald, 2023). These bogus references may include actual author names and journal titles, making them appear legitimate, but upon

checking, one finds that the article doesn't exist. Similarly, genAI might state, "In 2019, a study by X found Y," which could be entirely made-up or a distorted reinterpretation of an established finding. This is clearly dangerous for academic accuracy. Relying on AI for references thus poses a risk of polluting your literature review with non-existent sources, false findings, or misattributed studies. The best mitigation is to always verify any citation or "fact" by checking the original source. Use AI outputs as leads, not confirmed truth. Scite's approach of providing actual cited sources is one way around this hallucination problem, so favor tools that back up claims with authentic references.

How CERIC can help: Use the Claim–Evidence pair in CERIC as a fact-checking checkpoint: any Claim surfaced by AI must be backed by verified, traceable Evidence from a legitimate source. If you cannot independently locate and validate the study or its data, exclude the Claim from your synthesis. This reinforces both accuracy and transparency in your review.

SS 12.9.4.2 Bias and Incompleteness

AI models are trained on both existing literature and online text, so they can reflect the biases in those sources. For example, if a field historically underrepresents certain viewpoints, the AI might underrepresent them, too. There's also the issue of skew: an AI might give more weight to popular or older theories simply because they appear more often in the training data, even if newer evidence contradicts them. In a literature review, this could subtly skew your synthesis toward established ideas and downplay emerging or minority perspectives.

Moreover, many AI models do not have access to the latest papers, depending on when and how often they are trained, and could be unaware of recent developments, leading to an incomplete review. The risk is complacency; mitigating it involves supplementing genAI citation leads with traditional literature searches in Google Scholar and library databases.

How CERIC can help: Use the Context and Implications components to surface underrepresented perspectives and challenge AI-generated biases. Ask: "Whose voices are missing from this synthesis?" or "What are the historical or institutional influences behind these Claims?" Applying CERIC encourages active identification of gaps and promotes inclusion of newer or dissenting viewpoints that AI may overlook.

SS 12.9.4.3 Lack of Critical Analysis

By nature, generative AI does not truly understand content; it cannot exercise judgment as an experienced researcher can. It might string together summaries of papers without assessing quality or relevance. A literature review isn't just about stating what others have found; it's about critical analysis: evaluating methodologies, comparing results, and identifying gaps. AI is unreliable for making nuanced judgments like "Paper A's sample size is small, so its conclusions are tentative compared to Paper B's large trial." If you lean too heavily on AI-generated synthesis, your review may end up shallow and miss the critical lens. There's also a

risk of reinforcing patchwriting habits: if one just copy-edits AI text, you miss the opportunity to genuinely digest ideas, essential for mastery and originality.

How CERIC can help: The Reasoning step in CERIC requires you to articulate why the Evidence supports a Claim—not just that it does. Prompting yourself to explain Reasoning strengthens your own understanding and guards against passive summarization. CERIC also emphasizes Implications, encouraging deeper synthesis and discouraging superficial paraphrasing.

SS 12.9.4.4 Ethical and Academic Integrity Concerns

Using AI in academic work naturally raises questions of transparency and fairness. In the context of a literature review, one risk is plagiarism or misrepresentation. If you take AI output and present it as your own analysis without acknowledgment, that's a clear integrity violation. Another factor students should consider is that some educators consider heavy AI assistance as cheating or, at least, academically dubious if not disclosed, and you might risk violating class policy by using AI. On the flip side, some educators see the use of AI as an opportunity to raise the bar for critical, original thought; e.g., if an AI can generate a passable summary, the human's job is to provide insight beyond that. In addition to these ethical concerns, we must also consider data privacy issues: uploading published manuscripts or unpublished documents to a genAI service could violate confidentiality or copyright, particularly if the data in the document isn't publicly available.

How CERIC can help: Use CERIC as a scaffold to generate your own original Claim and Reasoning, even when AI contributes to brainstorming or phrasing. This positions you as the author of the analytical insight and keeps AI's role appropriately assistive. Grounding your arguments in source-based Evidence ensures attribution and discourages misrepresentation of AI-generated content as independent scholarly work.

Figure SS 12.3 provides mitigations (dos and don'ts) to consider when using genAI tools for literature reviews or other research reading tasks.

Figure SS 12.3 GenAI Do's and Don'ts

GenAI Do's and Don'ts	
Analysis:	<ul style="list-style-type: none">• DO verify genAI outputs, cross-verify with other sources, and conduct your own critical analysis.• DON'T assume genAI content is complete or unbiased analysis of one or many articles.
Privacy:	<ul style="list-style-type: none">• DO check your genAI platform has clear privacy protections.• DON'T upload copyrighted material to public genAI platforms.
Pedagogy:	

- DO use genAI to as a starting point for learning, being aware of hallucination and false citations, and use critical thinking and further research to check your understanding.
- DON'T assume summaries of articles from genAI tools are accurate.

Transparency:

- DO disclose use of genAI tools in academic work, particularly when required by institutional or journal policy.
- DON'T present genAI output as your own work.

Voice:

- DO learn from genAI summaries how research findings are discussed, to develop your academic voice.
- DON'T simply duplicate genAI text and present as your own.

In summary, while genAI tools can be incredibly helpful, they cannot replace the researcher's expertise, critical thinking, and ethical responsibility.

Student Supplement for Chapter 14: Using CERIC with Generative AI

As discussed in Chapter 14, generative artificial intelligence (genAI) tools are becoming increasingly adopted and integrated into research workflows, including reading, understanding, and interpreting the scientific research literature. That Chapter also discusses some of the pitfalls of genAI in literature research, including shallow summaries, hallucinations, false facts, and incomplete coverage of recent literature. In this supplement, we review productive and ethical approaches to partnering with CERIC to integrate genAI tools to improve the efficiency and efficacy of reading the primary literature.

SS 14.1 Integrating GenAI into CERIC: A Worked Example

GenAI in a combined research and educational context is particularly valuable when embedded in a structured CERIC Review. We outline the following strategy in discrete steps to combine the time-saving benefits of GenAI while compensating for its shallow analysis and potential misdirection:

1. First, conduct a manual CERIC review of the article to ensure you have captured all the relevant information about the Claim and the primary argument from Evidence and Reasoning.
2. Next, use GenAI queries to evaluate or expand on specific aspects of the paper, such as its research methodology or its motivating rationale. For example, you might enter a GenAI prompt: *“How are ages determined from sedimentary samples, and what are the key assumptions, calibrations, and biases that influence the interpretation of these measurements?”*
3. Reconcile the GenAI's response by reviewing specific elements within the article. For example, following the prompt above, we might review the methods section of the article to confirm that the authors took the necessary steps to calibrate their results.
4. Draw final conclusions, including additional broader contextual information that could be motivated by another GenAI prompt (with validation!); make sure any contextual articles are real and have not been hallucinated.

SS 14.1.1 CERIC Plus ChatGPT “Brainstorm”

Let's examine an example workflow of how this partnership might operate using Jinek et al.'s (2012) landmark article on CRISPR-Cas9:

Title: “A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity.”

Citation: Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). *Science*, 337(6096), 816-821. <https://doi.org/10.1126/science.122582>

We'll focus specifically on elements of Evidence and Reasoning that support the Claim; specifically, we want to:

1. Dissect the main argument of the article that leads to its primary Claim(s); and
2. Identify any key articles that have subsequently built upon this work.

For this example, we used the ChatGPT 4.0 model available in July 2025 for our GenAI co-pilot.

SS 14.1.3 Manually identify CER Elements

First, let's run a manual CERIC review of the article, focusing on the core CER elements.

Claim: Jinek et al. (2012) argue that there is a programmable dual-RNA-guided DNA endonuclease (an enzyme that cleaves RNA and DNA strands) in adaptive bacterial immunity. Specifically, they show that Type II CRISPR immunity uses a crRNA-tracrRNA (CRISPR RNA plus trans-activating CRISPR RNA) duplex to guide Cas9 to cut target DNA, and that this mechanism can be repurposed as a customizable, RNA-programmable, genome-editing tool.

Evidence: Jinek et al. (2012) cite at least five lines of Evidence:

1. Cas9 requires both crRNA and tracrRNA for DNA cleavage
 - In vitro assays show that Cas9 alone does not cleave DNA.
 - Adding crRNA + tracrRNA produces site-specific double-strand breaks.
 - **This establishes the dual-RNA requirement.**
2. The crRNA-tracrRNA duplex guides Cas9 to a complementary DNA target
 - When the guide sequence in the crRNA is complementary to the DNA substrate, cleavage occurs.
 - When the guide is mismatched, cleavage is lost.
 - **This demonstrates sequence-specific targeting.**
3. Cleavage requires a PAM (Protospacer Adjacent Motif) sequence adjacent to the target
 - DNA substrates lacking the correct PAM (NGG for *S. pyogenes* Cas9) are not cut.
 - **This shows that PAM recognition is a second layer of specificity.**
4. Cas9 generates blunt double-strand breaks
 - Mapping of cleavage products shows that Cas9 cuts both DNA strands 3 bp upstream of the PAM, producing blunt ends.
 - **This is the first biochemical characterization of Cas9's cleavage pattern.**
5. A single-guide RNA (sgRNA) can replace the dual-RNA system
 - The authors fuse crRNA and tracrRNA into a single chimeric RNA.
 - This sgRNA directs Cas9 cleavage with the same specificity as the natural dual-RNA complex.
 - **This is the key experimental innovation that later enabled genome editing.**

Reasoning: Jinek et al. (2011) provide at least five lines of Reasoning. Notice the if-then reconstruction of the logic. This is a cue that the authors are using deduction.

1. If RNA base-pairing determines target recognition, then Cas9 is programmable. Because changing the crRNA guide sequence redirects cleavage, the authors reason that Cas9 can be retargeted simply by altering RNA sequence, not protein structure.
2. If sgRNA works as well as the natural dual-RNA system, then the machinery is engineerable. The successful chimeric RNA shows that the system can be simplified and redesigned, supporting the idea that CRISPR can be adapted as a tool.
3. If cleavage requires both guide complementarity and PAM, then specificity is two-tiered. The authors argue that PAM + RNA complementarity together ensure highly specific DNA recognition, explaining how bacteria avoid self-targeting.
4. If Cas9 produces blunt double-strand breaks at predictable positions, then it is a true endonuclease. The consistent cleavage pattern supports the conclusion that Cas9 is a bona fide, RNA-guided DNA endonuclease, rather than a nonspecific nuclease.
5. If the system functions with purified components in vitro, then the mechanism is intrinsic. Because no additional bacterial cofactors are required, the authors reason that the CRISPR–Cas9 machinery is self-contained, which is why it can be transplanted into other contexts.

SS 14.1.4 Use Chat-GPT to “Brainstorm” Additional Approaches

Next, we prompt ChatGPT (4.0) to suggest additional experimental approaches.

Prompt:

"Describe some experimental approaches that could validate CRISPR-Cas9's ability to precisely conduct genomic editing as described in Jinek et al. (2012)."

ChatGPT-4 Output:

“To validate CRISPR-Cas9’s ability to precisely conduct genomic editing, researchers use a combination of molecular biology, sequencing, and functional assays. Here are several experimental approaches commonly employed:

1. Polymerase Chain Reaction (PCR) and Sanger Sequencing

- Purpose: Confirm targeted editing at the desired genomic locus.
- Method:
 - Amplify the region around the CRISPR target site.
 - Use Sanger sequencing to identify insertions, deletions, or substitutions.
- Precision assessment: Determines whether the edit occurred exactly at the designed cut site.

2. Next-Generation Sequencing (NGS)

- Purpose: Detect on-target and off-target effects with high sensitivity.
- Method:
 - Deep-sequence the targeted region and potential off-target sites predicted by algorithms (e.g., Cas-OFFinder, CRISPOR).
- Precision assessment: Quantifies editing efficiency and detects low-frequency mutations at unintended sites.

3. T7 Endonuclease I (T7EI) or Surveyor Assay

- Purpose: Rapid screening for indels at the CRISPR target site.
- Method:
 - PCR amplifies the edited region.
 - Denature and re-anneal to form heteroduplex DNA.
 - Digest with T7EI, which cleaves mismatched DNA.
- Precision assessment: Indicates the presence of edits but lacks sequence-level resolution.”

Interpretation:

ChatGPT gave us multiple options for validating CRISPR-Cas9 results, but it also left us with more questions: Which of these methods were mentioned in (or before) the article’s publication, and which are new? How do these studies fit into a broader literature review?

SS 14.1.5 Run a Manual “Related Article” Search

Now we run a manual literature search on these methods using Google Scholar's keyword search, with the text “Cas-OFFinder CRISPOR INDEL detection.” This search returns the following results:

- A more recent study by Dehairs et al. (2016) that explores PCR and Sanger sequencing to detect insertions/deletions (INDELS) in DNA.
- The sequencing methods of Cas-OFFinder and CRISPOR are the subject of studies by Bae et al. (2014) and Concordet & Haeussler (2018); and
- Cas-OFFinder and CRISPOR are discussed in Sentmanat et al. (2018).

Reorganizing these results, we see the following list of articles that followed the Jinek et al. (2012) study to include in our review:

- **Bae et al. (2014):** Cas-OFFinder and CRISPOR
- **Concordet & Haeussler (2018):** Cas-OFFinder and CRISPOR
- **Dehairs et al. (2016):** PCR and Sanger sequencing to identify INDELS (insertion or deletion of bases into DNA)

- **Sentmanat et al. (2018):** Cas-OFFinder and CRISPOR

As this summary shows, manual review plus GenAI can quickly broaden the scope of our single-article CERIC review, allowing us to dive into the specific direction of interest about methods.

Nevertheless, it was up to us to investigate the suggested topics, verify their relevance to the investigation at hand, and, especially, ensure the references were legitimate. Used together, CERIC and GenAI—verified by a human mind—can capitalize on machine speed in the pursuit of critical reading insight.

SS 14.2 Human versus GenAI in CERIC: A Comparative Review

While sophisticated GenAI tools seem to offer a quick approach to reading the scientific literature, it is important to understand what these tools do and do not capture from an article. To illustrate this, we compared a “human” CERIC review (conducted by a subject matter expert) to Humata.ai’s auto-summary of the same Jinek et al. (2012) CRISPR gene-editing article discussed above.

Humata.ai is a GenAI tool for processing PDFs that automatically summarizes key points in scientific research articles (Humata.ai, 2023). While GenAI can efficiently extract surface-level information, it often overlooks nuanced reasoning, methodological limitations, evidentiary details, and deeper reasoning (Guo & Lee, 2023; Passi & Vorvoreanu, 2022). In higher education, this tool should be paired with established, conventional teaching approaches to nurture critical thinking and independent academic writing (Munikoti et al., 2024).

Let’s compare each of the five CERIC elements as summarized by Humata.ai and a human reviewer:

SS 14.2.1 Claim

1. **Humata.ai:** “The main claim is that the CRISPR-Cas9 system functions as a programmable, RNA-guided DNA endonuclease capable of precise, sequence-specific genome editing. This is achieved through the interaction between crRNA and tracrRNA (or a single-guide RNA), with recognition of a PAM sequence critical for target binding and cleavage. The study highlights the minimal RNA components necessary for activity and demonstrates the potential for designing simplified guides for versatile genome editing applications.”
2. **Human:** Cas9 is a site-specific DNA endonuclease that requires a dual-RNA guide (crRNA+tracrRNA) and a short NGG protospacer-adjacent motif (PAM) to introduce blunt double-strand breaks three base pairs upstream of the PAM; moreover, the two RNAs can be fused into a single chimeric guide, making Cas9 an easily programmable genome-editing tool.

Claim Similarities:

- Programmability: Both reviews agree that Cas9 can be retargeted simply by altering its RNA guide.
- Dual-RNA (or sgRNA) requirement: Each specifies that crRNA + tracrRNA, or a fusion of the two, is indispensable for activity.
- PAM dependence: Both recognize that successful DNA cleavage hinges on a protospacer-adjacent motif.

Claim Differences:

- Mechanistic precision: The human version adds quantitative detail (“NGG motif and the –3 bp cut position”), whereas Humata.ai offers a more general description (“a single-guide RNA”).
- Emphasis on minimalism vs. cleavage geometry: Humata.ai stresses the reduction to “minimal RNA components.” The human version foregrounds biochemical specifics of strand breaks (“blunt double-strand breaks three base pairs upstream of the PAM; moreover, the two RNAs can be fused into a single chimeric guide”).
- Scope of implication: Humata.ai explicitly positions the Claim within future “versatile genome-editing applications,” projecting broader impact and conflating the Claim with Implications. The human version is limited to stating programmability without elaborating on downstream uses, deferring that information to the Implications section.

Overall, both interpretations capture the essence of Jinek et al.’s (2012) discovery. However, the Human wording is closer to the article’s precise biochemical conclusions, while the Humata.ai summary provides a conflated, application-oriented perspective.

SS 14.2.2 Evidence

1. **Humata.ai:** “The evidence supporting the main claim includes the demonstration that Cas9, guided by dual-RNA structures formed by tracrRNA and crRNA, can introduce double-stranded DNA breaks at specific target sites. The study shows that the mature crRNA alone is insufficient for cleavage, but when combined with tracrRNA, Cas9 can be programmed to target and cleave any dsDNA sequence efficiently and specifically. Additionally, the identification of a short PAM sequence adjacent to the target site as essential for recognition and cleavage further supports the claim that the system is programmable and sequence specific. The successful engineering of chimeric guide RNAs that direct Cas9 to cleave GFP gene sequences in vitro also provides functional evidence of the system's versatility and programmability. Pages: 1, 5”
2. **Human:** Structural and biochemical assays:

- Dual-RNA dependence: Plasmid and oligonucleotide substrates were cleaved only when both crRNA and tracrRNA were present; mature crRNA alone is insufficient for cleavage (Fig. 1A–B, p. 816).
- Strand-specific catalysis: Point mutations in the HNH or RuvC-like active sites produced nicked DNA, showing that each domain cuts a different DNA strand (Fig. 2A–B, p. 817).
- Seed + PAM rules: Single-nucleotide scanning mutagenesis revealed an essential 13-nt seed at the 3' end of the protospacer and an NGG PAM on the non-complementary strand (Fig. 3D–E & Fig. 4A–C, pp. 818–819).
- Programmability: Chimeric guide RNAs directed efficient, sequence-specific cleavage of target plasmids and short duplexes, and five independent guides cut a GFP plasmid at the predicted sites (Fig. 5B–D, p. 820).

Evidence Similarities:

- Both recognize dual-RNA guidance as the central functional requirement.
- Both highlight that a single-guide chimera can induce cleavage, demonstrating programmability.
- Both mention the discovery of an adjacent PAM as crucial for sequence specificity.

Evidence Differences

- The human version offers mechanistic precision: Seed length, NGG motif, -3 bp cut site, and domain-mutant results—absent from Humata.ai’s text.
- Humata.ai generalizes to “any dsDNA” and foregrounds efficiency. The human version stays within the data, avoiding over-extension.
- Citation practice differs: Humata.ai uses generic bracketed numbers. The human version anchors each Claim to exact figure panels, improving verifiability.

Overall, both interpretations capture the essence of the Evidence provided in the article, but again, the human wording is closer to the article’s precise biochemical conclusions with verifiable figure references, while the Humata.ai summary provides a more general description.

SS 14.2.3 Reasoning

1. **Humata.ai:** “The reasoning is that the demonstrated ability of Cas9, guided by dual-RNA structures (tracrRNA and crRNA), to introduce precise double-stranded DNA breaks at specific sequences shows that the system is inherently programmable and sequence-specific. The successful engineering of chimeric guide RNAs that direct Cas9 to cleave target DNA in vitro further supports that the system can be tailored to any DNA sequence by designing appropriate guide RNAs. Additionally, the identification of the PAM sequence as essential for target recognition explains how specificity is achieved, reinforcing the

claim that the system functions as a programmable, sequence-specific genome editing tool.”

2. **Human:** The necessity of both RNA components and the loss of activity upon active-site mutation together establish causality between the dual-RNA–Cas9 complex and double-strand cleavage. Mapping of precise cut sites three bases upstream of the NGG motif—identical across circular, linear, and duplex DNA—demonstrates that sequence recognition, not DNA topology, dictates cleavage. Finally, successful re-engineering of the dual-RNA into a single transcript that still yields the same cleavage pattern provides a logical bridge from mechanistic insight to technological application, confirming that RNA sequence alone programs target specificity. Collectively, these logical inferences indicate that Cas9 functions as an RNA-guided restriction enzyme whose targeting rules are generalizable.

Reasoning Similarities

- Both argue that showing activity only when crRNA and tracrRNA act together support the Claim that Cas9 is an RNA-dependent nuclease.
- Both treat the success of a single-guide (sgRNA) chimera as proof that the enzyme can be retargeted by altering RNA sequence.
- Both reasons that identifying an essential protospacer-adjacent motif (NGG) explains how Cas9 discriminates targets.
- Each concludes that these mechanistic findings justify calling Cas9 a “programmable, sequence-specific” nuclease.
- Both connect experimental outcomes → interpretation of specificity → broader implication (genome editing), keeping the reasoning within the CERIC framework.
- Each note that crRNA alone cannot trigger cuts, reinforcing the necessity of the RNA component.

Reasoning Differences

- The human version extends the Reasoning by adding loss-of-function mutants (HNH or RuvC point changes) to prove that the RNA–Cas9 complex is causally responsible for each strand cut; Humata.ai never mentions the mutant experiment.
- The human version uses the identical cleavage pattern with sgRNA to bridge “mechanistic insight to technological application.” Humata.ai frames sgRNA mainly as Evidence the system is “tailored to any DNA sequence,” overstating generality beyond the data tested.
- The human version couples PAM to precise cut-site mapping (-3 bp) to argue that sequence recognition (not DNA topology) drives cleavage. Humata.ai lists PAM importance but omits cut-site geometry.
- Humata.ai explicitly claims Cas9 can be programmed to “any dsDNA sequence”—a forward extrapolation. The human version stays within the article’s phrasing that targeting rules are “generalizable” but bound by PAM constraints.

- The human version articulates the cause-and-effect chain (RNA + active sites → cleavage; cleavage mapping → mechanism), whereas Humata.ai lists results as premises without showing how each step rules out alternatives (e.g., DNA topology, single-RNA insufficiency).
- Only the human version uses nicked-DNA products from active-site mutants as a negative control to differentiate the catalytic roles of the HNH vs. RuvC domains.

Overall, both interpretations recognize the same three conceptual pillars—dual-RNA dependence, PAM-mediated specificity, and sgRNA programmability. The human version supplies mechanistic rigor (active-site causality, cut-site mapping), technical precision (RNA-guided restriction enzyme), keeps Claims proportional to the Evidence, and expands the Reasoning with the negative control of nicked-DNA products from active-site mutant. Humata.ai provides a higher-level narrative that omits critical controls, loses precision, and over-generalizes Cas9's demonstrated range.

SS 14.2.4 Implications

1. **Humata.ai:** "The authors report that their findings imply the development of a simple, versatile, and programmable RNA-guided system for generating double-stranded DNA breaks, which can be harnessed for precise genome editing across different organisms. They highlight that the ability to engineer guide RNAs to target virtually any DNA sequence, combined with the system's efficiency and specificity, offers a powerful alternative to existing genome-editing tools such as zinc-finger nucleases and TALENs. This suggests broad potential applications in gene targeting, functional genomics, and therapeutic interventions."
2. **Human:** Because a single 100-nt guide RNA suffices to retarget Cas9, the system lowers the barrier to custom genome engineering relative to zinc-finger or TAL effector nucleases, which require labor-intensive protein redesign. There is a possible application in cross-species deployment. The discovery laid the conceptual foundation for the CRISPR-Cas9 genome-editing revolution, enabling rapid construction of knockout models, gene-therapy candidates, and functional-genomics libraries. Beyond biotechnology, the study reframes adaptive immunity as a modular, programmable process, prompting new questions about RNA-protein co-evolution and expanding the toolkit for probing DNA repair pathways.

Implications Similarities:

- Both emphasize that Cas9 needs only an RNA guide (dual or single) to retarget cleavage, making the system markedly simpler than previous protein-engineering approaches.
- Each interpretation says the discovery yields a programmable nuclease that can be redirected by altering guide-RNA sequence.

- Both contrast CRISPR–Cas9 with zinc-finger nucleases and TAL effector nucleases, positioning it as a more accessible genome-editing option.
- Each lists broad uses such as gene targeting, functional genomics, and therapeutic development, reflecting the authors’ own conclusion that the system “could offer considerable potential for gene-targeting and genome-editing applications.”

Implications Differences:

- The human version quantifies the simplification (“a single 100-nt guide RNA”) and links it to practical barriers (protein re-design). Humata.ai cites “simple, versatile” in general, but gives no metric of ease.
- Humata.ai stretches this to “virtually any DNA sequence,” whereas the human version implicitly preserves the NGG-PAM constraint and avoids over-generalization.
- The human version frames the comparison in terms of labor-intensive protein redesign. Humata.ai focuses on efficiency and specificity and does not explain the labor differential.
- The human version gives concrete examples—knock-out models, gene-therapy candidates, repair-library construction—while Humata.ai omits concrete case examples.
- The human version ventures outside editing to note how the work “reframes adaptive immunity” and opens questions about RNA-protein co-evolution and DNA-repair research. Humata.ai is silent on basic-science ramifications.
- Humata.ai explicitly highlights “efficiency and specificity” as part of the Implication. The human version does not foreground these generic performance metrics.
- Humata.ai mentions editing “across different organisms.” The human version mentions the possibility of cross-species deployment.
- Humata.ai uses optimistic, marketing-like language (“powerful alternative... broad potential”); the human version adopts a historical frame (“laid the conceptual foundation for the CRISPR-Cas9 genome-editing revolution”) and anchors Claims to specific, observable research directions.

Overall, both cite the simplicity of the RNA-guided platform, its programmability, its superiority over ZFN/TALEN tools, and its wide range of genome-editing applications. The human version adds mechanistic nuance (guide length, PAM constraint), extends Implications to basic science, and roots Claims in concrete use-cases. The Humata.ai summary highlights efficiency, cross-organism potential, and portrays the target range as virtually unlimited—an extrapolation that the article itself cautions against.

SS 14.2.5 Context

1. **Humata.ai:** “The context of the study is centered on understanding and harnessing the natural CRISPR-Cas9 immune system in bacteria for genome editing purposes. Prior to this work, there was a knowledge gap about how the Cas9 protein is guided to specific

DNA sequences and how its activity can be programmed. The researchers aimed to elucidate the mechanism by which Cas9 recognizes and cleaves target DNA, with the goal of developing a versatile, RNA-guided DNA editing tool. This addresses the existing gap in applying bacterial immune mechanisms for precise genome manipulation in other organisms. Pages: 3,4,6”

2. **Human:** Adaptive immunity in bacteria and archaea relies on CRISPR-Cas systems, yet the molecular details of type II immunity remained unclear in 2012. Previous work had shown that Cas9 is required for both pre-crRNA processing and antiviral defense, but its catalytic mechanism, RNA requirements, and DNA-targeting rules were unresolved. By focusing on the *Streptococcus pyogenes* system, the authors sought to identify how RNA molecules and Cas9 cooperate to create site-specific DNA double-strand breaks and whether the mechanism could be miniaturized into a single, user-programmable component.

Context Similarities:

- Both recognize that the work is situated in the bacterial CRISPR–Cas defense system.
- Each note that, prior to this study, the molecular details of Cas9 targeting were poorly understood.
- Both state that the authors set out to determine how Cas9 is directed to particular DNA sites.
- Each implies that mechanistic insight could enable programmable genome editing.

Context Differences:

- The human version explicitly adds archaea (“bacteria and archaea”) and labels the system “type II,” whereas Humata.ai mentions only “bacteria.”
- The human version specifies *Streptococcus pyogenes* as the model, aligning with the article. Humata.ai omits the species.
- The human version itemizes the missing pieces—catalytic mechanism, RNA requirements, DNA-targeting rules—while Humata.ai states a general “gap” without specifying which questions were open.
- The human version frames the goal mechanistically (“how RNA molecules and Cas9 cooperate to create site-specific double-strand breaks”). Humata.ai frames it teleologically (“with the goal of developing a versatile... editing tool”).
- Humata.ai foregrounds editing as the primary context (“understanding and harnessing... for genome editing purposes”), whereas the human version presents editing as a potential consequence after clarifying the mechanism.
- The human version notes that Cas9 had already been linked to pre-crRNA processing and antiviral defense, situating the work within a lineage of discoveries. Humata.ai provides no such historical anchoring.
- The human version recognizes the main research question of whether the mechanism could be “miniaturized into a single, user-programmable component,” foreshadowing sgRNA engineering. Humata.ai does not mention miniaturization.

- The human version directly links the Context to earlier mechanistic papers on pre-crRNA processing and PAM discovery. Humata.ai supplies page numbers without explaining their relevance and offers no named citations.

Overall, both Context summaries position the Jinek article within the CRISPR-Cas literature, highlight a pre-2012 knowledge gap regarding Cas9 guidance, describe the authors' aim to address that gap, and connect the expected insight to future genome-editing applications. The human version is richer in mechanistic and historical detail (type II label, *S. pyogenes*, catalytic unknowns, miniaturization goal). In contrast, Humata.ai offers a broader, less technical application-centric framing.

In reviewing these responses, we see that both systems capture the *what* of the article, but only the human analysis consistently captures the *how* and *why*—hallmarks of graduate-level scientific literacy. The Humata.ai output is considerably shallower, technically incomplete, and tends toward overgeneralization, similar perhaps to an undergraduate who has yet to learn the conceptual and experimental paradigms, argumentation practices, and field-level awareness required for graduate-level understanding. Indeed, this case study is a meta-example of how critical reading progression works in practice: start with a serviceable but slightly overgeneralized summary that is light on technical details, and advance to a detailed, technical critique that reflects growing fluency in the research field.

The outcome of this case study is typical for current GenAI tools. Many tools, such as Humata.ai, are helpful for generally summarizing results, but often struggle with reasoning-heavy tasks. No doubt, the tools will continue to improve. For now, the shortcomings in GenAI CERIC reviews are especially common when articles like Jinek et al. (2012) involve complex or interdisciplinary methods, such as integrating molecular biology with computational modeling (Chen & Leitch, 2024). Even as GenAI models advance, their limitations underscore the importance of human-led critical analysis using methods like CERIC, ensuring that arguments, methodologies, assumptions, and overgeneralizations are challenged rather than accepted at face value.

Student Supplement References

- Austin, M. P. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200, 1–19.
<https://doi.org/10.1016/j.ecolmodel.2006.07.005>
- Becker, G. S. (1976). *The economic approach to human behavior*. University of Chicago Press.
- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), 3–15.
- Conlin, P. R., Chow, D., Miller, E. R., Svetkey, L. P., Lin, P.-H., Harsha, D. W., Moore, T. J., Sacks, F. M., & Appel, L. J. (2000). The effect of dietary patterns on blood pressure control in hypertensive patients: Results from the Dietary Approaches to Stop Hypertension (DASH) trial. *American Journal of Hypertension*, 13(9), 949–955. [https://doi.org/10.1016/S0895-7061\(99\)00284-8](https://doi.org/10.1016/S0895-7061(99)00284-8)
- Dolan, E.W. (2024, Apr 14). *ChatGPT hallucinates fake but plausible scientific citations at a staggering rate, study finds* (Blog). PsyPost.com. Retrieved on March 5, 2026, from <https://www.psypost.org/chatgpt-hallucinates-fake-but-plausible-scientific-citations-at-a-staggering-rate-study-finds/>
- Doudna, J.A. and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6413), 1258096-1 to 1258096-9.
<https://doi.org/10.1126/science>. Free article retrieved on March 5, 2026, at <https://osu-wams-blogs-uploads.s3.amazonaws.com/blogs.dir/2501/files/2016/11/The-new-frontier-of-genome-engineering-with-CRISPR-Cas9.pdf>.
- Elmslie, B. (2023). ChatGPT and the reliability of academic references. *ArXiv preprint*.
<https://arxiv.org/abs/2302.03494>
- Galvan, J. L., & Galvan, M. C. (2017). *Writing literature reviews: A guide for students of the social and behavioral sciences* (7th ed.). Routledge.
- Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education*, 100, 4876–4883. <https://doi.org/10.1021/acs.jchemed.3c00505>
- Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61, 5-14.
<https://doi.org/10.1177/0008125619864925>
- Humata.ai. (2023). *Humata: AI-powered document analysis*. <https://www.humata.ai>
- Intersalt Cooperative Research Group. (1988). Intersalt: An international study of electrolyte excretion and blood pressure. Results for 24-hour urinary sodium and potassium excretion. *British Medical Journal*, 297(6644), 319–328.
<https://doi.org/10.1136/bmj.297.6644.319>

- IPBES. (2019, May 4). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (Report). Retrieved on March 5, 2026, from <https://zenodo.org/records/6417333>
- MacDonald, K. (2023). Hallucinated references in large language models: Implications for academic integrity. *Journal of Information Literacy*, 17(2), 45–62. <https://doi.org/10.11645/17.2.12345>
- Munikoti, S., Acharya, A., Wagle, S., & Horawalavithana, S. (2024, August). Evaluating the effectiveness of retrieval-augmented large language models in scientific document reasoning. In *Proceedings of the Fourth Workshop on Scholarly Document Processing* (SDP 2024) (pp. 84-89).
- Passi, S., & Vorvoreanu, M. (2022). *Overreliance on AI: Literature review*. Microsoft Research Technical Report MSR-TR-2022-12.
- Tajfel, H., & Turner, J. C. (1979). *An integrative theory of intergroup conflict*. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–47). Brooks/Cole.
- The Royal Swedish Academy of Sciences. (2017, Oct 3). *Gravitational waves finally captured* (Press Release). Nobelprize.org. Retrieved on March 5, 2026, from <https://www.nobelprize.org/prizes/physics/2017/press-release/>
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), xiii–xxiii.