



Using Text Mining to Identify Teleological Explanations in Physics and Biology Textbooks: An Exploratory Study

Richard Brock¹ · Nikos Tsourakis² · Kostas Kampourakis³

Accepted: 1 March 2024
© The Author(s) 2024

Abstract

Creating and critiquing explanations of phenomena is a significant goal of many scientific disciplines and therefore also a learning goal of science education. A significant source of explanations is science textbooks; however, the large corpus of text in textbooks means that manual review of explanations by individual researchers is extremely time consuming. In this paper, we introduce a text-mining approach for identifying legitimate and illegitimate forms of teleological explanations in school physics and biology textbooks. An ongoing debate exists about the legitimacy of teleological explanations, that is, explanations which account for a phenomenon by reference to a final end, purpose, or goal. Until recently, researchers tended to view teleology as an illegitimate form of scientific explanation. Recent theoretical cases in biology and physics have emphasized that legitimate teleological explanations exist in both domains. Eight science textbooks used in England and internationally were analyzed for instances of teleological explanations. The analysis reveals the efficiency of the text-mining approach for automating the analysis of textbooks and its potential as a research approach in science education. In considering text mining as a research approach, we report terms that are likely to be associated with legitimate teleological explanations. We found that legitimate teleological accounts are used by textbook authors, and we present novel categorizations of these forms of explanation. We argue that text mining can be a useful approach in science education research and our findings suggest guidance for both textbook writers and teachers related to their selection of legitimate explanatory forms.

Keywords History and philosophy of science · Language of science and classrooms · Philosophy of science · Science education · Textbooks

✉ Richard Brock
Richard.brock@kcl.ac.uk

¹ School of Education, Communication & Society, King's College London, London, UK

² Department of Translation Technology, University of Geneva, Geneva, Switzerland

³ Section of Biology and IUFE, University of Geneva, Geneva, Switzerland

1 Introduction

Teleological explanations are causal explanations that refer to a purpose or a goal (Ariew, 2007; Brock & Kampourakis, 2023; Kampourakis, 2020; Lennox, 1992; Walsh, 2008). For instance, when we say that a person is going to the bookstore in order to buy a book, we are explaining their actions in terms of an intention. Or when we say that animals have hearts in order to pump blood, we explain the presence of hearts in terms of their function in the body. When students say that organisms change their features in order to adapt to their environments, they explain the changes they observe in terms of a perceived need. Finally, we can say that we design airplanes with long wings and powerful engines in order for them to be able to fly, describing the design underlying their features. What matters is that in all of the aforementioned explanations, there is a common feature: something exists or happens in order for something else to occur. All such accounts are teleological explanations of different kinds, but not all of them are scientifically legitimate.

Research in psychology has found that teleological explanations are used in different contexts by people of different cultural backgrounds (e.g., Kelemen, 1999; Kelemen et al., 2005; Kelemen et al., 2013; Rottman et al., 2017; Schachner et al., 2017). Therefore, teleological explanations seem to be due to a cognitive bias that emerges in childhood and persists into adulthood, and so, it is important to consider the legitimacy of teleological explanations in teaching. Most importantly, teleological explanations are not necessarily wrong. However, the adjective “teleological” has been extensively used in the science education literature to describe students’ misconceptions, especially in biology (e.g., Kampourakis & Zogza, 2008; Lennox & Kampourakis, 2013; Trommler & Hammann, 2020). Therefore, it is useful to consider which teleological explanations are legitimate and which are not, rather than reject them altogether.

Teleological explanations describing the intentions of conscious agents are perfectly legitimate. So, when we say that we go to the bookstore in order to buy a book, or that we design airplanes with specific features in order for them to be able to fly, we describe our own intentions to achieve a goal, and therefore, we explain why we did whatever we did in order to achieve that goal. These teleological explanations based on intentions and design are legitimate as an account of our behavior or of the construction or use of artifacts. However, it has been found in research that these kinds of explanations are often extrapolated to nature, because people—especially young children—perceive the features of organisms and of natural objects in the same way that they perceive the features of artifacts (Kelemen, 2012).

Teleological explanations are quite prevalent in biology where they are often indicated by phrases such as “in order to,” “so that,” and “for the sake of.” Whereas in the past teleological explanations were considered to be illegitimate in biology, philosophers have long shown that some kinds of teleological explanations in biology can be legitimate (e.g., Lennox, 1993; Brandon, 1981). A careful analysis of teleological explanations shows that there are different kinds, with some being based on design and intentions and others being based on natural selection (Lennox & Kampourakis, 2013). When it comes to science teaching and learning, legitimacy is determined not by the structure of the explanation itself but by the underlying causal account; that is, legitimacy depends on whether the causes referred to match causes in the world. Legitimate teleological explanations reflect underlying causal structures, even if the causes are not referred to directly. By contrast, illegitimate teleological explanations cite as causes entities that do not bring about the phenomena of interest, or in some cases even reverse causality. Teleological explanations based on design or need

are considered illegitimate in biology, whereas those based on natural selection are considered to be legitimate. So, among the examples mentioned above, it is legitimate to say that animals have hearts in order to pump blood, because we thus explain the presence of hearts in terms of their function in the body. In short, something exists because of what it does, and so, we can say that it exists for doing it. In contrast, saying that organisms change their features in order to adapt to their environments is an illegitimate teleological explanation because it accounts for the changes observed in terms of a perceived need (Kampourakis, 2020).

By contrast to biology, teleological explanations are a relatively understudied topic in physics (Brock & Kampourakis, 2023). Therefore, in the present paper, we contrast teleological explanations in biology and physics textbooks. The large quantity of text in the sample required a systematic and exhaustive analysis, so we adopted text-mining methods. Our analysis is based on a corpus consisting of eight English science textbooks. The English context was chosen because it is a country where national examination boards work with publishers to produce textbooks which are widely used in schools. The widespread use of standard texts means that the explanations in the examination board books potentially influence a large cohort of students' understanding of scientific ideas and are, therefore, worthy of study. The search was carried out by defining a list of phrases that marked potential teleological statements. Our intention from this analysis is not to generalize about the prevalence of forms of teleological explanations in science textbooks but to identify patterns in a sample of legitimate and illegitimate teleological explanations and to reflect on the usefulness of the text-mining approach in the context of science education research.

2 Theoretical Background

2.1 Legitimate and Illegitimate Teleological Explanations in Physics

Whilst their legitimacy in the context of biology education has been acknowledged (Kampourakis, 2020; Lennox & Kampourakis, 2013), until recently, discussion has suggested that teleological explanations in the context of physics education are likely to be illegitimate (Trommler & Hammann, 2020). In physics, the entities of interest are often inanimate objects or abstract concepts which do not have agency, and therefore, teleological explanations implying their intentions are often considered illegitimate. However, we have recently argued that there are cases of legitimate teleological explanations in physics education (Brock & Kampourakis, 2023). Our argument draws on Lange's (2017) model of constraint-based teleology. Lange observes that some explanations are legitimate even when they do not specify a causal agent. Legitimate causal explanations can highlight constraints that make an outcome inevitable, without directly referring to a cause. Consider the claim that a "... compact star will shrink to minimize the total energy, eventually collapsing to a black hole" (Schaffner-Bielich, 2020, pp. 86–87). The active cause in the star's collapse, gravitational force, is not referred to. Instead, the author cites a universal constraint, the conservation of energy, to indicate a necessary final state, the reduction in total energy of the system. In legitimate constraint teleological explanations, the end state required by a constraint can be implied by referring to the constraint itself. For example, in proposing the existence of a force, "... a force called radiation-reaction force, must be present in order to avoid violating the conservation law of energy" (Cornille, 2003, p. 388), an end state in which the final energy of a system is equal to its initial value is implied. Lange (2017)

argues that constraints of this form can indicate which outcomes are inevitable and which are impossible and are a legitimate explanatory form.

A condition on constraints is that they must be axiomatic and invariant across contexts (Lange, 2017). We therefore propose that teleological explanations that draw on laws, such as the law of conservation of energy in the example above, are legitimate because they rely on universal constraints. By contrast, teleological explanations that invoke rules or laws that are only true in some situations and lack the invariance required of constraints are not legitimate. For example, consider the explanation of a rotating object: "...for there to be a resultant force towards the centre..., the frictional force must increase" (OCR, 2019, p. 28). In this case, the change in frictional force is explained by the need for a centripetal force, a condition that is only found in rotating systems. The requirement for a resultant force acting towards the center of rotation is not a necessary condition; it is not present in all systems; hence, this form of teleology, we argue, is illegitimate. Some examples of legitimate and illegitimate teleological explanations (Brock & Kampourakis, 2023), with an indication of the reason for their legitimacy, are shown in Table 1.

Searching for the examples in Table 1 to support our argument for the existence of legitimate teleological explanations in physics, raised the question of how frequent such explanations were in science textbooks. Table 1 has been generated in an unsystematic fashion, from spontaneous notes from our reading and from an unstructured search. To develop a robust analysis of the frequency of different types of teleological explanations, we developed a text-mining approach.

2.2 Text Mining as a Means to Analyze a Large Corpus

Human language is a structured communication system based on grammar and vocabulary, with the distinctive feature of being *compositional*. We can combine or rearrange words to form new sentences with little effort. It is also *referential*, allowing us to refer to people, objects, or events from the past or future. To understand human language, however, we must overcome several challenges related to its peculiarities (Tsourakis, 2022). For instance, when we write, we tend to omit a lot of common-sense knowledge, assuming that the reader possesses it. The inherent ambiguity of natural language cannot be resolved without the proper context. Consider, for example, the word *break*, which can be interpreted as a pause from doing something, but it can also refer to a personal or social separation. Besides lexical ambiguity, we can encounter syntactic ambiguity, as in the following phrase: *The fish is ready to eat*. Is the fish ready to be fed, or can we eat the fish now? Similarly, in the case of identifying irony and sarcasm, positive or negative words can connote the opposite of their normal meaning, such as *yeah*, *sure*. Even worse, human texts can contain stereotypes and biases that prohibit their use in systems for the general public. Finally, when dealing with languages that lack the necessary linguistic resources, such as data, tools, and language technologies, we may face constraints that hinder our ability to analyze relevant material.

Linguistics is the main field for studying human languages and applying scientific methods to questions about their nature and function (Department of Linguistics, The Ohio State University, 2022). In recent decades, computational approaches to linguistic questions have gained ground, giving rise to text mining. The latter defines the processes of extracting useful and relevant information from unstructured textual data (Tsourakis, 2022). Text mining's ability to identify patterns, relationships, and insights within this data source makes it an attractive alternative to manual human scrutiny for quickly processing large quantities of

Table 1 A summary of the application of teleological explanation in the context of physics (from Broek & Kampourakis, 2023)

| Example | Legitimate (L) or illegitimate (I) | Reason |
|--|------------------------------------|---|
| a) "... the compact star will shrink to minimize the total energy, eventually collapsing to a black hole" (Schaffner-Bielich, 2020, pp. 86–87) | L | The causal networks invoked are appropriate and the principle of minimization of total energy is invariant across contexts |
| b) "In order for momentum to be conserved, the products of the reaction must be moving and, therefore, they must have some kinetic energy" (Christodoulides, 2016, p. 233) | L | Appropriate cause and effect is implied and conservation principles have a high degree of invariance |
| c) "... a force called radiation-reaction force, must be present in order to avoid violating the conservation law of energy" (Cornille, 2003, p. 388) | L | The prevailing causal relationships are reflected, and a universal principle is cited |
| d) "...an antibaryon must appear simultaneously to keep the baryon number the same" (Motz & Weaver, 2013, p. 339) | L | The existing causal structure is invoked, and conservation laws have appropriate necessity |
| e) "When asked why a balloon is round, we might say it chooses this form over others to equalize interior pressure" (Schoemaker, 1991, p. 210) | I | This account is illegitimate due to the anthropomorphic implication and the insufficiently necessary constraint (equalization of pressure) |
| f) "...a gas expands with increasing T[emperature] in order to keep P[ressure] constant" (Jespersen & Hystlop, 2022, p. 520) | I | Whilst the cause-and-effect relationships are appropriately reflected, the constraint draws on a ceteris paribus law and hence is insufficiently necessary |
| g) [In the context of a rotating object]: "...for there to be a resultant force towards the centre the frictional force must increase" (OCR, 2019, p. 28) | I | A legitimate causal relationship is referred to but, because the presence of a resultant force is not a requirement across contexts, the necessity of the constraint is too limited |

text. It also has the advantage of being less biased than manual methods as it relies on data-driven insights rather than human intuition or subjectivity. With the exponential growth of digital content, text-mining techniques allow for a more comprehensive understanding of the science education texts, including teaching materials and research papers. For instance, researchers might identify the underlying themes or topics within a large corpus, find terms and concepts of interest, extract the attitudes and opinions of writers in a particular area, and so forth (e.g., Byeon et al., 2021). Conversely, text mining can lack contextual understanding and, thus, lead to incorrect interpretation of nuances in a corpus. Often, domain-specific knowledge is needed to analyze scientific texts accurately, meaning algorithmic techniques can misclassify phrases. Having humans involved in the analysis process provides a means to interpret textual data categorized automatically, generating insights that can lead to informed decisions. We use the latter, hybrid, analytical approach in this paper.

2.3 Text Mining in Science Education Research

A recent systematic review searched publications for studies that used text mining in the context of science and mathematics education (Shin & Shim, 2021). Shin and Shim identified 64 articles that applied the technique between 2010 and 2019, 41 in the context of science education. The two most common uses for the technique in science education were for modeling students' cognition by analyzing their responses to probes for research purposes and in automated assessment. The authors found only two studies which used text mining for the same end as us, documentary analysis. Wahlberg and Gericke (2018) used text mining to analyze how protein synthesis was described in Swedish secondary textbooks. Reitsma et al. (2012) used text mining to examine the alignment of US curriculum documents from different standard authoring bodies. In addition to the research identified by Shin and Shim (2021), one additional study applied text-mining techniques in the context of science education. Jiang and McComas (2014) analyzed popular science texts for the inclusion of content related to the nature of science. In reflecting on the application of text mining to the context of science education, Jiang and McComas (2014) concluded:

The successful application of the text mining technique in the current study opened a new branch for science education research, which invites more applications of such technique on the analysis of other aspects of science textbooks, popular science writing, or any other materials involved in science teaching and learning. (p. 1804)

Our study extends the so far limited application of text mining in science education research to the examination of teleology and considers the affordances and limitations of the technique. We ask two questions: In what contexts do physics and biology textbook authors make use of legitimate and illegitimate teleological explanations? To what extent is text mining a useful approach for identifying instances of legitimate and illegitimate teleological explanations in school science textbooks?

3 Methods

3.1 Textbook Sample

We approached publishers of school science textbooks in England with a request for electronic copies of textbooks used to teach physics and biology curricula to 11–18-year-old

students. In response, we received a set of eight textbooks suitable for the analysis from a single publisher (Table 2). The books form a convenience sample—our intention is not to draw general conclusions about textbook authors’ use of teleology; rather, we aim to report usage in the case of books from one publisher and to reflect on the potential of the text-mining approach. The texts included biology and physics textbooks focused on preparation for external examinations at age 16 (AQA Physics and Biology - GCSE) and for examination at age 18 (AQA Biology and Physics A-level and Advanced Biology/Physics for You). In the English system, in many schools, one teacher (who may be a specialist in biology, chemistry, or physics) teaches content across all three disciplines to students aged 11–14 years old. The sample included two textbooks (Activate 1 and 2) with biology, chemistry, and physics content aimed at 11–14-year-old students. Given that our focus is on the legitimacy of some forms of teleology in biology and physics, the analysis mainly focuses on text from the biology and physics books, with a smaller number of sentences drawn from the general science books (which include biology and physics topics).

3.2 Using Text Mining to Identify Teleological Explanations

For this exploratory study, we analyzed the convenience sample of three physics, three biology, and two science textbooks (see Table 2), using the AntConc corpus analysis software (Anthony, 2022). Researchers and linguists use this software to analyze and explore large corpora of texts, as it supports several languages and can handle texts in various formats, including plain text, HTML, XML, and PDF. In our case, all textbooks were available in PDF format. After specifying a list of pertinent terms for teleology, we loaded the whole corpus into the tool. Then, we performed a Key-Word-In-Context (KWIC) search for each term, providing concordance results. The search results display a keyword surrounded by a context of several words on either side. This technique allows us to see how words and phrases are commonly used in a corpus of texts. Initially, the phrases we looked for were the following: “in order to”; “in order that”; “for the sake of”; “to this end”; “for this reason”; “to achieve this”; “and so”; “as”; “as a result of”; “because”; “due to”; “hence”; “since”; “so that”; “therefore”; and “thus.” We used the maximum context size offered by the tool, equal to 25 tokens on the left of the search term and 25 tokens on the right. In addition, the software provides color highlighting for the term and the surrounding text to enhance readability. All results were saved into an Excel file for further processing.

Table 2 Summary of the eight texts analyzed (*B*, biology; *P*, physics; *S*, science. Note, in the paper, the books are referred to by their codes, B1, B2, etc.)

| Code | Book | Year of publication | Target age group |
|------|---|---------------------|------------------|
| B1 | AQA Biology - A-level | 2015 | 16–18 |
| B2 | AQA Biology - GCSE | 2016 | 14–16 |
| B3 | Advanced Biology for You | 2015 | 16–18 |
| P1 | AQA Physics - A-level | 2015 | 16–18 |
| P2 | AQA Physics - GCSE | 2015 | 14–16 |
| P3 | Advanced Physics for You | 2015 | 16–18 |
| S1 | Activate 1 (Biology, Physics and Chemistry) | 2017 | 11–14 |
| S2 | Activate 2 (Biology, Physics and Chemistry) | 2017 | 11–14 |

In the next step, the first coder examined the terms. The excerpts surrounding the search phrases were classified into the categories of legitimate teleology, illegitimate teleology, not teleological, and unclear by the coder. We used dropdown menus in Excel with predefined options based on the previous categories to facilitate the annotation task. A second coder then checked the first classification and disagreements were resolved through discussion. To limit the responses in the tables below, we set a criterion of removing terms for which 5% or lower of the responses were categorized as legitimate teleology in either the biology or physics books.

4 Results

Our results report data related to the frequency of use of different forms of teleology in the sample textbooks. We first report counts of types of teleological explanations in the physics textbooks in our sample, consider the challenges of categorization and the limitations of text mining, and conclude with a comparison between teleological explanations in the physics and biology textbooks we studied.

4.1 The Prevalence of Teleological Explanations in Physics Textbooks in the Sample

We searched all the sample texts for the phrases mentioned in Sect. 3. The following four phrases were not found in any of the physics textbooks: “in order that”; “for the sake of”; “to this end”; “for this reason.” The number of instances of the other terms that met the 5% criterion for inclusion discussed above and the number of legitimate and illegitimate teleological explanations among them are presented in Table 3. As shown there, except for the phrase “so that” and “in order to,” in all other cases, the number of legitimate teleological explanations is relatively low. It is also interesting that, overall, the number of legitimate teleological explanations is always higher than the number of the illegitimate ones. Many of these legitimate cases relate not to natural phenomena, but to human intentions or intentional behavior. Therefore, even though the number of legitimate teleological explanations is not high overall, it is nevertheless important that legitimate teleological explanations are used in textbooks. As teleology has often been associated with misconceptions, in our view, the kinds of legitimate teleological explanations that exist in textbooks merit consideration.

Table 3 indicates that, in our sample of physics books, three terms were particularly likely to be associated with legitimate teleology. These are “to achieve this” (4/4 cases), “in

Table 3 The number of incidences of phrases in the physics texts (P1, P2, P3) indicating teleological explanations, categorized as legitimate or illegitimate, unclear, and not teleological

| Term | Total number | Legitimate teleology | Illegitimate teleology | Not teleological | Unclear |
|-----------------|--------------|----------------------|------------------------|------------------|---------|
| aim(s) to | 2 | 0 | 0 | 2 | 0 |
| in order to | 36 | 35 | 0 | 0 | 1 |
| so that | 199 | 142 | 8 | 47 | 2 |
| to achieve this | 4 | 4 | 0 | 0 | 0 |
| want(s) to | 29 | 4 | 0 | 25 | 0 |

order to” (35/36 cases), and “so that” (142/199 cases). No terms were particularly likely to be associated with illegitimate teleology. We attempt to explain the pattern by considering subcategories of teleology. The examples of legitimate teleology in physics textbooks were coded into the subcategories of constraint and intentional teleology (discussed below) by two coders. Illustrative examples of legitimate and illegitimate forms of these two subcategories, drawn from all search terms, are shown in Table 4.

The examples in Table 4 include two categories of legitimate teleological explanations in physics, legitimate intentional teleology, and legitimate constraint-based teleology. The examples of legitimate intentional teleology in the table (related to vehicles (the “hence” example) and electrical plugs (using “so that”)) refer to the intentions of the designers of objects. Elsewhere (Brock & Kampourakis, 2023), we have argued that an additional form of legitimate teleological explanations in physics arises when constraints are cited as causes. For example, the entry in Table 4 for the term “because” is “The ball loses an equal amount of momentum, because the total momentum is conserved.” The law of conservation of momentum is given as the cause of the ball’s change in momentum, an argument from a constraint. Constraint-based teleology is premised on the assumption that constraints have sufficient necessity; that is, they are invariant across contexts and over time (Lange, 2017). The principle of conservation of momentum in this case is an example of a constraint due to its invariance. Relationships that do not have a sufficient degree of necessity, we have suggested, cannot be legitimately cited as causes. For example, an illegitimate example in Table 4 refers to Ohm’s law: “The p.d. V must also fall exponentially. $V=IR$, and so I is proportional to V . \therefore The current I must fall exponentially” (P5, p. 307). Ohm’s law is an example of *ceteris paribus* law (Cartwright, 1980), that is, not a true law because the relationship between the variables only holds if some condition is met (temperature remains constant). We have argued that such cases are illegitimate because there are contexts in which the relationship is not true. Amongst the cases of legitimate teleology in physics textbooks, we identified two categories of legitimate teleology: cases that refer to an agent as a cause, and cases that cite a constraint. In total, in the physics books, we found 171 cases of legitimate constraint teleology and 66 of legitimate intentional teleology. Table 3 includes an interesting pattern. Some phrases are more likely to indicate legitimate teleological explanations than others. Phrases that met the inclusion criterion and were only found in legitimate uses in the physics context were “in order to,” “to achieve this,” and “want(s) to” (Table 3). After coding cases as legitimate or illegitimate, the cases were then coded in categories by type of explanation (Table 5).

For our sample of physics textbooks, some terms are more likely to be linked to one category of legitimate teleology. Consider the case of a term that was cut from the main analysis because of its low instances of legitimate cases, “due to.” Both legitimate instances of “Due to” are examples of constraint (“Due to conservation of energy, ...” (P2, p. 12) and “...due to the invariance of the speed of light” (P3, p. 501)). In the cases of “in order to,” in physics contexts, 34 out of 35 of the responses were coded as legitimate agential teleology (for example, “the active [radioactive] wastes are dealt with in order to reduce the hazards” (P1, p. 487)). It is interesting to observe that “due to” is more likely to indicate a constraint, by comparison with “in order to” which tends to indicate human action. Science education textbook writers might also note that the terms “as” (6% of cases categorized as illegitimate), “so that” (6%) and “because” (5%) had relatively high proportions of instances in which their use was categorized as illegitimate (though the total counts for “as” and “because” are relatively low). For example, in the case of “as,” a force, friction, is implied to have a goal: “As it is trying to stop things moving, friction always acts parallel to the surfaces in contact and in the” (P3, p. 20). When coding, we found that in the cases

Table 4 Illustrative examples of subcategories of legitimate and illegitimate teleological explanations for different terms in the physics textbooks. Bold text added for emphasis

| Phrase | Legitimate teleological explanation | Illegitimate teleological explanation |
|-----------|--|--|
| And so | <i>Legitimate constraint teleology</i> “No electrons are lost from the circuit, and so the number of coulombs passing any point each second must be the same conventional current” (P5, p. 244) | <i>Illegitimate constraint teleology</i> “The p.d. V must also fall exponentially. $V = I R$, and so I is proportional to V . ∴ The current I must fall exponentially” (P5, p. 307) |
| As | <i>Legitimate constraint teleology</i> “Lenz’s law states that the direction of the induced current is always such as to oppose the change that causes the current” (P1, p. 413) | <i>Illegitimate intentional teleology</i> “Friction tries to prevent motion between two surfaces in contact. As it is trying to stop things moving friction always acts parallel to the surfaces in contact” (P5, p. 20) |
| So that | <i>Legitimate intentional teleology</i> “The inside of the case is shaped so that the wires and the pins can’t touch each other when the plug is sealed.” (P2, p. 66) | <i>Illegitimate intentional teleology</i> “When the waves reach your ears, they make your eardrums vibrate in and out so that you hear sound.” (P2, p.180) |
| Therefore | <i>Legitimate constraint teleology</i> “After one loop of the circuit, the energy it has gained must be equal to the energy it has dissipated. Therefore : Around any closed loop in a circuit, the sum of the e.m.f.s is equal to the ...” (P5, p. 251) | No illegitimate teleological explanation |

Table 5 Subcategories of legitimate teleological explanations in physics textbooks by term

| Term | Total legitimate teleology | Legitimate constraint teleology | Legitimate intentional teleology |
|--------------------|----------------------------|---------------------------------|----------------------------------|
| aim to | 0 | 0 | 0 |
| in order to / that | 35 | 1 | 34 |
| so that | 142 | 138 | 4 |
| to achieve this | 4 | 0 | 4 |
| want(s) to | 4 | 0 | 4 |

of “in order to” and “to achieve this,” the inter-rater agreement was high (97.3% and 100%, respectively) and the likelihood of teleological legitimacy was similarly high (97.3% and 100%, respectively). This agreement can be explained by data in Table 5—both “In order to” and “to achieve this” are very likely to indicate human agency, making them relatively uncontroversial to classify. Two terms that were particularly likely to be associated with legitimate teleology “to achieve this” (4/4 cases) and “in order to” (35/36 cases) are both terms in which the majority of instances are agential. The relative easiness of identifying human action may explain the high proportion of legitimate application.

4.2 Some Cases of Disagreement in Coding

The use of text mining might imply that instances of particular forms of explanation can be easily searched for, by their association with particular phrases (like those in Table 3). However, our dual coding process emphasizes that the relationship between words and meaning is complex, and raters differed in their interpretation of whether particular claims were teleological or not. Ambiguity in language, subjectivity, and contextual factors contribute to these disagreements (Aroyo & Welty, 2015). Ambiguity arises from multiple meanings and unclear references, subjectivity stems from different perspectives and biases, and contextual factors vary based on cultural background and domain-specific knowledge. For example, in the case of “so that,” the coders were in agreement in about 69% of the cases. This is a relatively low percentage of agreement (an analysis of thresholds of coder agreement suggests that 90% agreement is accepted by all and 80% agreement by many researchers (Neuendorf, 2002)), and normally, the coders should discuss their criteria and recode anew. In contrast, the inter-coder agreement for “in order to” and “to achieve this” were 97.3% and 100%, respectively, and in general, there was high inter-coder agreement. In cases where there was disagreement, we discussed the cases and, in most instances, agreed on the categorization of the disciplinary expert coder. However, we have decided to report the results of this initial coding of phrases including “so that” here (see Table 6), because we found some interesting and consistent patterns of disagreement.

The disagreements between coders, in the examples above, can be categorized into two groups, which can be thought of as operating in a hierarchy. First, different interpretations of terms determined whether an excerpt was categorized as teleological or not. The conjunction “so that” can be read with two meanings, the first indicating a purpose (The man ate the cake so that his hunger was satisfied) and second indicating an end state, but one that is not causal, where “so that” is read to mean “in such a way as to” (The leaf fell and landed so that it was perfectly covering the stone). In the case of the final example in Table 6 “The block swings **so that** the centre of mass rises a vertical

Table 6 Selected examples of coding differences for “so that” in physics textbooks. Bold text added for emphasis. L = Legitimate, I = Illegitimate, NT = Not teleological

| Excerpt | Coder 1 | Coder 2 |
|--|---------|---------|
| “The free electrons in each sphere now spread out, so that the charge is distributed evenly over the surface of each sphere” (P5, p. 289) | I | L |
| “Drag and weight may eventually balance so that the projectile falls at its terminal velocity” (P5, p. 48) | I | L |
| “When the waves reach your ears, they make your eardrums vibrate in and out so that you hear sound” (P2, p. 180) | I | L |
| “The light reflects/transmits off an object that is luminous/non-luminous into your eye so that you see it.” (P3, p. 137) | I | L |
| “The total p.d. across both lamps is 6 V. This is shared between the two lamps, so that each lamp has a p.d. of 3 V across it” (P5, p. 245) | I | L |
| “Usually, the protons are orientated randomly, so that their magnetic moments cancel out.” (P5, p. 438) | NT | I |
| “The field around the spire is strong enough to ionise the air, so that charge can flow between the cloud and the spire.” (P5, p. 295) | I | L |
| “The block swings so that the centre of mass rises a vertical distance of 0.15 m” (P5, p. 121) | NT | L |

distance of 0.15 m” (P5, p. 121 emphasis added), the sentence can be read as a claim that the swing occurs in order for the center of mass to reach a particular height. Alternatively, the “so that” clause can be taken to mean “in such a way that”; that is, the sentence means the mass swings in such a way that the center of mass rises 0.15 m, or simply “The block swings **so** the center of mass rises a vertical distance of 0.15 m.” Strictly, the meaning of “so that” indicates purpose (Collins, n.d.) and the sentence should be categorized as a teleological case. It can be further categorized as illegitimate (if the intention is ascribed to the block) or legitimate, if agency is attributed to the person who swung the object (but who doesn’t appear in the sentence). In this case, it might be assumed that the author did not intend to ascribe agency to an inanimate object.

Second, constraints are defined as having necessity; that is, they apply over time and across contexts (Lange, 2017). For an explanation to be categorized as legitimate constraint teleology, the conditions cited must have the property of necessity. However, the threshold of necessity required for a condition to be judged as a constraint is unclear. Consider the sentence: “The free electrons in each sphere now spread out, **so that** the charge is distributed evenly over the surface of each sphere” (P5, p. 289, emphasis added). The end state in this case is a uniform distribution of charge. The conditions that cause that distribution can be interpreted in different ways. First, the uniform distribution of charge can be seen as arising from a variational principle, whereby systems tend to an end state with a local minimum of electrical potential energy. Variational principles related to the conservation of energy can be seen as legitimate constraints (Brock & Kampourakis, 2023). Alternatively, the end state of uniform distribution might be interpreted as contingent on the absence of other local charges, and hence, the condition can be interpreted as lacking the necessity expected of a constraint. Detecting cases of legitimate constraint teleology in the context of physics can be subtle and open to interpretation. For example, in the case of “The total p.d. across both lamps is 6 V. This is shared between the two lamps, **so that** each lamp has a p.d. of 3 V across it” (P5, p. 245, emphasis added). One interpretation is that the sentence implies some agency in the circuit; that is, there is an intention behind the distribution of potential difference, which would be illegitimate. By contrast, the claim can be seen as a form of constraint teleology, and the potential differences across the bulbs must add to the supply EMF, due to conservation of energy, a legitimate constraint. We chose to code the sentence as legitimate constraint teleology. In the context of a falling object, the sentence “Drag and weight may eventually balance **so that** the projectile falls at its terminal velocity” (P5, p. 48, emphasis added) was interpreted as illegitimate by the first coder as it implies an inanimate object, the projectile, has an intention, and legitimate by the second coder as an instance of constraint teleology. On discussion, it was agreed that whilst the reading of “so that” indicates a condition linked to Newton’s first law of motion, that is, no net force leads to motion at constant velocity, the constraint is not necessary, as there can be periods of acceleration and motion at constant velocity depending on the conditions. The case was agreed as an instance of illegitimate teleology.

Two cases in Table 6 refer to the human senses and allow an application of a category of legitimate teleology from the biological context—that of functions. The disciplinary expertise (or lack thereof) of the coders may underlie the disagreement in this case. Coder 1 categorized both examples as illegitimate, due to the implication that waves travel with the intention of being perceived.

When the waves reach your ears, they make your eardrums vibrate in and out **so that** you hear sound. (P2, p.180)

The light reflects/transmits off an object that is luminous/non-luminous into your eye **so that** you see it. (P3, p. 137)

An alternative reading of these cases, preferred by coder 2 and agreed upon, is that they are cases of biological functional teleology, a legitimate form of teleology (discussed above). Such differences in coding suggest that the categorization of some statements can require discussion and may not be easily automated, as we discuss in the conclusion.

4.3 Comparison to the Respective Findings in Biology Textbooks

Teleology in biology is a topic that has been widely researched and has been considered a conceptual obstacle to understanding evolution (e.g., Kampourakis & Zogza, 2009) and heredity (e.g., Stern et al., 2022). However, as explained in Sect. 1, the problem is not teleology and teleological explanations per se, but the underlying design stance (Kampourakis, 2020). Students can legitimately explain that something exists for a purpose or role, insofar as this purpose or role has emerged (evolved) through natural processes such as natural selection. So, the key question is not whether teleological explanations are used in biology textbooks, but rather whether these are legitimate or not. In biology, illegitimate teleological explanations are those that are based on design, intentionality for non-conscious entities, or other anthropomorphic accounts. For instance, a rabbit can run and hide in order to avoid a predator—this is a behavior (intentional or instinctive, it does not make much difference). However, a rabbit cannot change its color in order to conceal itself and hide from the predator, nor can a population of rabbits gradually change their color in order to conceal themselves, unless this occurs through a process of natural selection (there is variation, and some individuals have a survival and reproduction advantage over others).

As we argue above for the case of physics textbooks, in the three biology textbooks, we have limited the search to terms where the percentage of legitimate teleological explanations was over 5% across both books (see Table 7).

The sentences containing these phrases were coded independently by two coders, and in all cases, there was more than 95% agreement. The legitimate teleological explanations in Table 7 were classified into five different kinds of cause (see Table 8):

- Explanations related to human intention, that is, about something that was done (in nature or in the laboratory) in order for humans to achieve something.
- Explanations referring to a physiological function, that is, something happening within organisms—above the cellular level—that contributed to something or maintained their physiology.

Table 7 The number of incidences of phrases in the biology texts (B1, B2, B3) indicating teleological explanations, categorized as legitimate or illegitimate, not teleological, and unclear

| Phrase | Total number | Legitimate teleology | Illegitimate teleology | Not teleology | Unclear |
|-------------------|--------------|----------------------|------------------------|---------------|---------|
| aim(s) to | 6 | 4 | 0 | 2 | 0 |
| in order to/ that | 111 | 111 | 0 | 0 | 0 |
| so that | 150 | 130 | 4 | 12 | 4 |
| to achieve this | 9 | 9 | 0 | 0 | 0 |
| want(s) to | 39 | 5 | 0 | 31 | 3 |

Table 8 Subcategories of legitimate teleological explanations in biology textbooks by term

| Term | Total legitimate teleology | Legitimate: human intention | Legitimate: physiological function | Legitimate: molecular function | Legitimate: animal behavior | Legitimate: adaptation | Legitimate: other/unclear |
|-----------------|----------------------------|-----------------------------|------------------------------------|--------------------------------|-----------------------------|------------------------|---------------------------|
| aim(s) to | 4 | 3 | 1 | 0 | 0 | 0 | 0 |
| in order that | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| in order to | 110 | 50 | 30 | 18 | 3 | 9 | 0 |
| seek to | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| so that | 130 | 59 | 34 | 29 | 1 | 2 | 5 |
| to achieve this | 9 | 4 | 4 | 0 | 1 | 0 | 0 |
| want(s) to | 5 | 5 | 0 | 0 | 0 | 0 | 0 |

- Explanations related to a molecular function, that is, something happening within or around cells that contributed to something or maintained their physiology.
- Explanations citing an adaptation, in most cases with some form of the term “adapt-” itself included.
- Explanations referring to animal behavior.

Most legitimate teleological explanations were found to be related to the phrases “in order to” and “so that,” with an almost 100% agreement between the two coders. Let us consider the sentences that contained the phrase “in order to” (see Table 9, note there was one instance of “in order that”), for which all 110 sentences were coded as legitimate teleological:

- 50 referred to a human intention, that is, about something that was done (in nature or in the laboratory) in order for humans to achieve something.
- 30 referred to a physiological function, that is, something happening within organisms—above the cellular level—that contributed to something or maintained their physiology.
- 18 referred to a molecular function, that is, something happening within or around cells that contributed to something or maintained their physiology.

Table 9 Examples of subcategories of legitimate teleological explanations in biology textbooks containing the phrase “in order to.” Bold text added for emphasis

| Sentence containing the phrase “in order to” in legitimate teleological explanations | Subcategory of legitimate teleology |
|--|-------------------------------------|
| “Large areas of land may need to be cleared in order to produce biofuels.” (B3, p. 476) | Human intention |
| “Antigens are treated before being introduced to the body of the patient in order to make them relatively harmless.” (B3, p. 289) | Human intention |
| “...animals possess two principal coordinating systems: the nervous system, which communicates rapidly, and the hormonal system, which usually communicates more slowly. Both systems interact in order to maintain the constancy of the internal environment while also being responsive to changes in the external environment.” (B1, p. 386) | Physiological function |
| “The human eye is a complex organ containing photoreceptors, and able to control the amount of light that enters it, refract light rays in order to focus them, and transduce light energy into action potentials.” (B3, p. 353) | Physiological function |
| “This doubling of the number of chromosomes would continue at each generation. It follows that, in order to maintain a constant number of chromosomes in the adults of a species, the number of chromosomes must be halved at some stage.” (B1, p. 224) | Molecular function |
| “Others are carrier proteins that bind to ions or molecules like glucose and amino acids, then change shape in order to move these molecules across the membrane.” (B1, p. 84) | Molecular function |
| “Petals of insect-pollinated flowers usually have scent and make nectar in order to attract insects.” (B3, p. 116) | Adaptation |
| “It [a niche] includes all the biotic and abiotic conditions to which an organism is adapted in order to survive, reproduce and maintain a viable population.” (B1, p. 467) | Adaptation |
| “As reptiles have no physiological means of cooling, they must seek shade in order to reduce their body temperature.” (B1, p. 633) | Animal behavior |

- 9 referred to an adaptation, in most cases with some form of the term “adapt-” itself included.
- 3 referred to animal behavior.

The phrase “so that” was also related to legitimate teleological explanations, even though not all of them were legitimate. Overall, we found 150 sentences with “so that” of which 130 were legitimate teleological, 4 were illegitimate teleological, 12 were non-teleological, and 4 were unclear. Among the 130 legitimate teleological explanations, we found the following categories (Table 10).

- 59 referred to a human intention, that is, about something that was done (in nature or in the laboratory) in order for humans to achieve something (in several cases there were also instructions to students using the textbooks).
- 34 referred to a physiological function, that is, something happening within organisms - above the cellular level - that contributed to something or maintained their physiology.
- 29 referred to a molecular function, that is, something happening within or around cells that contributed to something or maintained their physiology.
- Four had unclear categorizations.
- Two referred to an adaptation, in one case with the term “adapt” itself included.
- One referred to an animal behavior.
- One referred to a constraint of the kind found in physics textbooks.

Table 10 Examples of subcategories of legitimate teleological explanations in biology textbooks containing the phrase “so that.” Bold text added for emphasis

| Sentence containing the phrase “so that” in legitimate teleological explanations | Subcategory of legitimate teleology |
|---|-------------------------------------|
| “Artificial hearts can also be used to give a diseased heart a rest, so that it can recover.” (B2, p. 59) | Human intention |
| “Scientists hope that eventually they will be able to genetically engineer faulty human pancreatic cells so that they work properly.” (B2, p. 165) | Human intention |
| “In bright light, the iris makes the pupil very small. This reduces the amount of light that goes into the eye, so that too much bright light does not damage the delicate, light-sensitive cells.” (B2, p. 154) | Physiological function |
| “Pollination is needed in order to bring the two male gametes (inside a pollen grain) near to the female gamete so that fertilisation can occur.” (B3, p. 117) | Physiological function |
| “During meiosis, four haploid cells are produced, each being genetically different. It is important that gametes are haploid so that at fertilisation a diploid zygote is formed.” (B1, p. 657) | Molecular function |
| “Binding of phosphate ion [sic] to protein causes the protein to change shape so that access for the molecules is open to the inside of the membrane but closed to the outside.” (B1, p.94) | Molecular function |
| “These bacteria have adaptations to their cytoplasm so that water does not move out of their cells into their salty environment by osmosis.” (B2, p. 269) | Adaptation |
| “In order to gain heat lizards orientate themselves so that the maximum surface area of their body is exposed to the warming rays of the Sun.” (B1, p. 380) | Animal behavior |
| “This means that when these polar phospholipid molecules are placed in water they position themselves so that the hydrophilic heads are as close to the water as possible and the hydrophobic tails are as far away from the water as possible...” (B1, p. 17) | Constraint |

Table 11 Examples of subcategories of illegitimate teleological explanations in biology textbooks containing the phrase “so that.” Bold text added for emphasis

| Sentences containing the phrase “so that” in illegitimate teleological explanations | Subcategory of illegitimate teleology |
|---|---------------------------------------|
| “Species X may change the environment so that it becomes more suitable for species Z.” (B1, p. 497) | Intentionality |
| “Crossing over: the process whereby a chromatid breaks during meiosis and rejoins to the chromatid of its homologous chromosome so that their alleles are exchanged.” (B1, p. 657) | Intentionality |
| “To obtain white offspring the green parent must provide a recessive allele so that a double recessive can occur in the offspring.” (B1, p. 637) | Intentionality |
| “Valves in the cardiovascular system are designed so that they open whenever the difference in blood pressure either side of them favours the movement of blood in the required direction.” (B1, p. 175) | Design |

Table 12 Summary of the counts and percentages of illegitimate and legitimate teleological explanations related to “in order to” and “so that” in disciplinary textbooks

| Phrase | Physics textbooks (P1, P2, P3) | | Biology textbooks (B1, B2, B3) | | Science textbooks (S1, S2) | |
|-------------|--------------------------------|---------------|--------------------------------|---------------|----------------------------|--------------|
| | L | I | L | I | L | I |
| in order to | 35/36 (97.2%) | 0/36 (0%) | 110/110 (100%) | 0/110 (0%) | 2/2 (100%) | 0/2 (0%) |
| so that | 142/199 (71.4%) | 8/199 (0.04%) | 130/150 (86.6%) | 4/150 (0.02%) | 17/24 (70.1%) | 1/24 (0.04%) |

Furthermore, among the four sentences in which instances of illegitimate teleology were found, three referred to some kind of intentionality and one to design. Cases of illegitimate teleology in the biology textbooks are presented in Table 11.

Given these results, it is now interesting to compare the findings in the physics and the biology textbooks. Table 12 presents an overview of the results that reveal some interesting similarities. We can see legitimate teleological explanations related to “in order to” and “so that,” the two most prevalent terms, occur with approximately the same frequency in both physics and biology textbooks.

The findings in Table 12 are interesting from both a theoretical and a methodological point of view. Having made a case for the legitimacy of teleological explanations in physics based on constraints, it is interesting to find that the sample physics textbooks include a significant number of legitimate explanations of this form. Within this small sample, markers of legitimate teleological seem to have differing prevalence in the biology and physics textbooks in the sample, with “in order to” being nearly three times as common in biology as in physics, perhaps related to the agential implication of the term. The more neutral “so that” which can be applied both to agents and constraints has a similar prevalence between the disciplines. In considering the value of text mining as a research approach, the table presents some grounds for optimism, at least in the context of searching for teleological explanations in textbooks. The term “in order to” is associated with a high probability in the physics, biology, and general science textbooks in the sample of identifying legitimate teleological explanations. Such findings may act as useful guidance and caveats to science education textbook writers, teachers, and editors. When writing a sentence with the phrases

“so that” and “in order to,” our findings can prompt an author that they may be using a teleological explanation of some form, and to consider the guidance we have discussed on the legitimacy of various uses. As such, our analysis has led to practically useful advice for those writing scientific explanations.

5 Conclusions

This exploratory study confirms that legitimate forms of teleological explanations are used by the science textbook authors in our samples, and the findings suggest several recommendations for teachers, science textbook writers, and researchers interested in using text-mining approaches. Before considering the implications of our findings, it is worth re-emphasizing the caveat that the conclusions are based on a small set of textbooks from one publisher. Therefore, we will avoid generalizations about the prevalence of legitimate and illegitimate teleological explanations in science textbooks in general. First, we note that, in the set of science textbooks analyzed, both legitimate and illegitimate biological and physical teleological explanations, of a number of types, were found. This suggests a simplistic categorization of teleology as illegitimate, or legitimate only in the contexts of agents, should be discarded. Our discussion of coding indicates judgements of the legitimacy of teleological explanations can be subtle and nuanced. When writing a sentence (or giving a verbal explanation in class) using one of the terms (e.g., “in order to”) linked to both legitimate and illegitimate uses, textbook authors and teachers should reflect on the conditions we suggest for legitimate use. In some cases, for example, in the context of judging whether a constraint is sufficiently necessary, judgements can be subjective and depend on semantic inferences (see the discussion above about the meaning of “so that”). We hope that the identification of markers that are likely to identify legitimate forms of teleological explanation reliably, across contexts (notably, “so that” and “in order to,” in the context of the books analyzed), can be replicated with larger data sets to provide general guidance to authors and teachers.

An aim of our study was to consider the potential of text mining as a research tool in science education. Our exploratory study highlights the potential of the tool, but also suggests that even when text searching is systematic, interpretations of language rely on subjective interpretations of meaning, so that coding text, at least in the case of teleological explanation, is also somewhat subjective. We found, during the coding process, that the context of phrases was significant for our judgements. For example, consider this fragment of a sentence: “...the area over which diffusion can take place is greatly increased. Therefore the rate of diffusion is also greatly increased, so that much more of a substance moves in a given time” (B2, p. 15). When considered in isolation, the sentence might be considered a candidate for a legitimate constraint teleological explanation—as the area of some surface is increased, the rate of diffusion across the surface will increase due to the rules of probability. However, by expanding the search, the full sentence is revealed to be:

By folding up the membrane of a cell, or the tissue lining an organ, the area over which diffusion can take place is greatly increased. Therefore the rate of diffusion is also greatly increased, so that much more of a substance moves in a given time. (B2, p. 15)

The reference to the function of the intestine suggests the explanation should be categorized as legitimate functional teleology. It is difficult to predict how much text might be required to

reliably categorize a fragment. Despite such uncertainties, there is scope for greater automation of the search process we used. For example, the phrase “in order to conserve” might be used to find cases of legitimate constraint teleology related to conservation laws (for example, “in order to conserve energy,” “in order to conserve momentum,” etc.).

Contemporary accounts suggest there is a lack of agreement over criteria for “good” scientific explanations (Alameh & Abd-El-Khalick, 2018). De Regt and Dieks (2005) and Alameh and Abd-El-Khalick (2018) propose a pragmatic approach to explanation that draws on a “toolbox” of different approaches in different contexts. Rather than expecting there to be definitive rules for categorizing legitimate from illegitimate explanatory approaches, we argue that there is likely to be some scope for interpretation in the classification of cases, and hence, a fully automated approach to text mining may not be possible. Whilst human coding of cases is responsive, it is also time consuming and inconsistent; by contrast, automated coding is rapid and reliable but lacks the flexibility of human intelligence (Heap et al., 2017). A hybrid mining approach (as in our study which used an automated initial search procedure to create a manageable data set for human coding) might have the benefits of both approaches. Text mining can, by reporting patterns of current usage, support the development of recommendations for textbook authors and teachers writing instructional material. In the case of teleology for example, our search suggests that authors and teachers, in the context of physics teleological explanations, are cautious when they use the phrases “and so,” “as,” “in order to,” and “therefore.” In our sample, these phrases were associated with relatively high probabilities of illegitimate uses of teleology (albeit the total incidence of such cases was low). In this way, text-mining approaches have the potential to offer guidance to writers of science education textbooks and other teaching resources about appropriate use of scientific language.

The data set used in this study is relatively small (eight texts), and future studies, by drawing on larger corpora, will be able to report generalizable claims about language use in science education texts. For example, it would be interesting to determine patterns in the prevalence of categories of legitimate teleological explanations (e.g., human intention, physiological function, molecular function) in a larger sample of textbooks. More generally, text mining could be used to determine authors’ explanatory approaches in different contexts. Given the suggestion that approaches to explanation are contextual (Alameh & Abd-El-Khalick, 2018; De Regt & Dieks, 2005) and not necessarily clearly determined by theoretical rules, empirical data on usage patterns (for example, what forms of explanation are common in different topics in disciplines?) will be useful data for authors and teachers. Data related to trends in usage over time would also be of interest. We encourage researchers in science education to adopt the text-mining approach—it has the potential to develop novel knowledge about the customary linguistic usage of a field and hence support clarity of communication in science education texts.

Acknowledgements We are grateful to the publisher of science textbooks who provided electronic copies of their texts for our analysis.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alameh, S., & Abd-El-Khalick, F. (2018). Towards a philosophically guided schema for studying scientific explanation in science education. *Science & Education*, 27, 831–861.
- Anthony, L. (2022). *AntConc* (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Ariew, A. (2007). Teleology. In D. Hull, & M. Ruse (Eds.), *Cambridge companion to the philosophy of biology* (pp. 160–181). Cambridge University Press.
- Aroyo, L., & Welty, C. (2015). Truth is a lie: Crowd Truth and the seven myths of human annotation. *AI Magazine*, 36(1), 15–24.
- Brock, R., & Kampourakis, K. (2023). A justification of legitimate teleological explanations in physics education: An argument from necessary constraints. *Science & Education*, 32(4), 927–945.
- Brandon, R. N. (1981). Biological teleology: Questions and explanations. *Studies in History and Philosophy of Science Part A*, 12(2), 91–105.
- Byeon, Y. J. J., Islamaj, R., Yeganova, L., Wilbur, W. J., Lu, Z., Brody, L. C., & Bonham, V. L. (2021). Evolving use of ancestry, ethnicity, and race in genetics research—A survey spanning seven decades. *The American Journal of Human Genetics*, 108(12), 2215–2223.
- Cartwright, N. (1980). The truth doesn't explain much. *American Philosophical Quarterly*, 17(2), 159–163.
- Christodoulides, C. (2016). *The Special Theory of Relativity. Foundations, Theory, Verification, Applications*. Springer International Publishing.
- Collins (n.d.). *Definition of 'so that'*. Online Dictionary. Retrieved January 25, 2024, from: <https://www.collinsdictionary.com/dictionary/english/so-that>.
- Cornille, P. (2003). *Advanced Electromagnetism and Vacuum Physics*. Worth Publishers, Inc.
- De Regt, H. W., & Dieks, D. (2005). A contextual approach to scientific understanding. *Synthese*, 144, 137–170.
- Department of Linguistics, The Ohio State University (2022). *Language Files: Materials for an Introduction to Language and Linguistics*. 13th edition. Ohio State University Press.
- Heap, B., Krzywicki, A., Schmeidl, S., Wobcke, W., & Bain, M. (2017). A joint human/machine process for coding events and conflict drivers. In G. Cong, W.-C. Peng, W. E. Zhang, C. Li, & A Sun (Eds.) *Advanced Data Mining and Applications: 13th International Conference, ADMA 2017, Singapore, November 5–6, 2017, Proceedings 13* (pp. 639–654). Springer International Publishing.
- Jespersen, N. D., & Hyslop, A. (2022). *Chemistry the Molecular Nature of Matter*. John Wiley and Sons, Inc. Wiley.
- Jiang, F., & McComas, W. F. (2014). Analysis of nature of science included in recent popular writing using text mining techniques. *Science & Education*, 23, 1785–1809.
- Kampourakis, K. (2020). Students' "teleological misconceptions" in evolution education: why the underlying design stance, not teleology per se, is the problem. *Evolution: Education and Outreach*, 13, 1–12.
- Kampourakis, K. (2022). Reconsidering the goals of evolution education: defining evolution and evolutionary literacy. *Evolution: Education and Outreach*, 15(1), 21.
- Kampourakis, K., & Zogza, V. (2008). Students' intuitive explanations of the causes of homologies and adaptations. *Science & Education*, 17(1), 27–47.
- Kampourakis, K., & Zogza, V. (2009). Preliminary evolutionary explanations: A basic framework for conceptual change and explanatory coherence in evolution. *Science & Education*, 18, 1313–1340.
- Kelemen, D. (1999). Function, goals and intention: Children's teleological reasoning about objects. *Trends in Cognitive Sciences*, 3(12), 461–468.
- Kelemen, D. (2012). Teleological minds: How natural intuitions about agency and purpose influence learning about evolution. In K. S. Rosengren, S. K. Brem, E. M. Evans, & G. M. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 66–92). Oxford University Press.
- Kelemen, D., Callanan, M. A., Casler, K., & Pérez-Granados, D. R. (2005). Why things happen: Teleological explanation in parent-child conversations. *Developmental Psychology*, 41(1), 251.
- Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, 142(4), 1074–1083.

- Lange, M. (2017). *Because Without Cause Non-Causal Explanations in Science and Mathematics*. Oxford University Press.
- Lennox, J. G. (1992). Teleology. In E. Lloyd, & E. F. Keller (Eds.), *Keywords in Evolutionary Biology* (pp. 122–127). Harvard University Press.
- Lennox, J. G. (1993). Darwin was a teleologist. *Biology and Philosophy*, 8, 409–421.
- Lennox, J. G., & Kampourakis, K. (2013). Biological teleology: the need for history. In K. Kampourakis (Ed.), *The philosophy of biology: A companion for educators* (pp. 421–454). Springer.
- Motz, L., & Weaver, J. H. (2013). *The story of physics*. Springer Science and Business Media B.V.
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Sage.
- OCR (2019). *Exemplar Candidate Work Physics A* Cambridge Assessment.
- Reitsma, R., Marshall, B., & Chart, T. (2012). Can intermediary-based science standards crosswalking work? Some evidence from mining the standard alignment tool (SAT). *Journal of the American Society for Information Science and Technology*, 63(9), 1843–1858.
- Rottman, J., Zhu, L., Wang, W., Schillaci, S., Clark, R., K. J., & Kelemen, D. (2017). Cultural influences on the teleological stance: Evidence from China. *Religion Brain & Behavior*, 7(1), 17–26.
- Schachner, A., Zhu, L., Li, J., & Kelemen, D. (2017). Is the bias for function-based explanations culturally universal? Children from China endorse teleological explanations of natural phenomena. *Journal of Experimental Child Psychology*, 157, 29–48.
- Schaffner-Bielich, J. (2020). *Compact star physics*. Cambridge University Press.
- Schoemaker, P. J. H. (1991). The quest for optimality: A positive heuristic of science? *Behavioral and Brain Sciences*, 14(2), 205–215.
- Shin, D., & Shim, J. (2021). A systematic review on data mining for mathematics and science education. *International Journal of Science and Mathematics Education*, 19, 639–659.
- Stern, F., Kampourakis, K., & Müller, A. (2023). “Genes for a role”, “genes as essences”: Secondary students’ explicit and implicit intuitions about genetic essentialism and teleology. *Journal of Research in Science Teaching*, 60(2), 237–267.
- Trommler, F., & Hammann, M. (2020). The relationship between biological function and teleology: Implications for biology education. *Evolution: Education and Outreach*, 13(11), 1–16.
- Tsourakis, N. (2022). *Machine learning techniques for text: Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation*. Packt Publishing Ltd.
- Wahlberg, S. J., & Gericke, N. M. (2018). Conceptual demography in upper secondary chemistry and biology textbooks’ descriptions of protein synthesis: A matter of context? *CBE Life Sciences Education*, 17(3), ar51.
- Walsh, D. (2008). Teleology. In M. Ruse (Ed.), *The Oxford handbook of philosophy of biology* (pp. 113–137). Oxford University Press.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.