



# The Development of Cognitive Load Theory: Replication Crises and Incorporation of Other Theories Can Lead to Theory Expansion

John Sweller<sup>1</sup> 

Accepted: 11 September 2023 / Published online: 19 September 2023  
© The Author(s) 2023

## Abstract

Cognitive load theory has been in development since the 1980s. Much of the impetus for that development has come from firstly, replication failures using randomised controlled trials and secondly, from the incorporation of other theories into cognitive load theory. Both have led to theory expansion. The immediate cause of the so-called “replication crisis” in psychology and other disciplines is a failure to replicate previous empirical findings. Using cognitive load theory as an example, I argue that the appearance of contradictory evidence does not necessarily derive from a failure to properly collect data. Rather, it can be caused by initially insufficiently detailed theories, with increasing detail often revealing the reason for a failure to replicate. For cognitive load theory, each failure to replicate, rather than being a negative, contributed to the further development of the theory. In addition, the theory has developed over many years by closely incorporating other theories associated with human cognitive architecture and evolutionary psychology. In this paper, I discuss some of the developmental milestones associated with cognitive load theory and how they were informed by replication failures and theory integration.

**Keywords** Cognitive load theory · Human cognitive architecture · Instructional design · Working memory · Long-term memory · Evolution by natural selection · Replication crisis

---

This article is part of the Topical Collection on Theory Development in Educational Psychology.

---

✉ John Sweller  
j.sweller@unsw.edu.au

<sup>1</sup> School of Education, University of New South Wales, Sydney, NSW 2052, Australia

## Introduction

The replication crisis in psychology (Open Science Collaboration, 2015) is rightly considered to have negative implications. Science cannot progress without replicable data. Nevertheless, using cognitive load theory as an example, I will argue in this paper that replication failures, along with the integration of compatible theories, can have positive effects on theory development.

Replication failure frequently is attributed to failures associated with the data collection and treatment process that can be due to a variety of factors such as p-hacking and post-hoc explanations. One corrective to such problems is to use the Open Science procedure of pre-registration of proposed studies (Gehlbach & Robinson, 2018, 2021), which can eliminate many of the issues associated with data collection and treatment. There may be another factor unrelated to the collection and treatment of data that is relevant to replication failure. That factor is inadequate theory development which can contribute to a particular category of replication failure associated with “conceptual replications” as opposed to “direct replications” (Plucker & Makel, 2021). Conceptual replications support a concept or general procedure but direct replication is an attempt to accurately replicate a study. Conceptual replication is usually based on a theory and its associated hypotheses. Irvine (2021, P. 845) indicated the distinction in the following manner: “Direct replications are those in which the (relevant aspects of) experimental procedures of a selected study are reproduced as closely as possible. Successful direct replications help to rule out false positives and possible experimenter effects. Conceptual replications retain the basic theoretical reasoning in the selected study but use different procedures or operationalisations of variables to test an experimental hypothesis. Successful conceptual replications provide information about the underlying theory’s “generalisability”. A failure in conceptual replication may be due to inadequate theory development (Greene, 2022; Oberauer & Lewandowsky, 2019).

In this paper, I argue that not only can inadequate theory development lead to conceptual replication failure, but conceptual replication failure can also be an important component of subsequent theory development. Direct replication failures are less likely to be due to inadequate theory development and more likely to be due to problems associated with experimental procedures such as a failure to accurately replicate, or variations in sample sizes. Cognitive load theory (Sweller et al., 2011, 2019), an instructional theory based on our knowledge of human cognitive architecture and evolutionary psychology, is used as a case study for the perspective that conceptual replication failure can be due to an inadequate theory and discovery of that inadequacy can lead to useful theory development. Multiple conceptual replication failures when compared with replication successes can be used as a major trigger to consider further theory development. In addition, the integration of compatible theories can provide another source of theory development. In this paper, I will outline the current version of cognitive load theory, its origins, and its development following replication failures and the incorporation of other theories.

## Current Version of Cognitive Load Theory

The influence of evolutionary psychology on the field's understanding of human cognitive architecture provides a base for cognitive load theory. The theory assumes the distinction between categories of knowledge made by Geary (2002, 2005, 2008, 2012; Geary & Berch, 2016) who distinguished between biologically primary and secondary knowledge. Biologically primary knowledge is knowledge humans have evolved to acquire over many generations. Examples are general problem-solving strategies or listening to and speaking one's native language. This frequently very complex knowledge is acquired easily, automatically, and unconsciously simply by immersion in a suitable environment. It normally does not need to be explicitly taught. Engaging in species-typical childhood behaviours involving interaction with other people or objects is all that is required.

In contrast, biologically secondary knowledge is knowledge that one's culture deems important. Humans have evolved to acquire this knowledge in a general sense, but unlike biologically primary knowledge, we have not evolved to acquire specific examples of secondary knowledge such as learning to read and write or learning mathematics. Virtually everything that is taught in educational contexts consists of biologically secondary knowledge. Educational institutions were developed to teach secondary knowledge because if it is not explicitly taught, it is less likely to be acquired (Hattie, 2008). It is generally not acquired easily, automatically, or unconsciously but rather needs conscious effort. Schools were invented to assist in the acquisition of biologically secondary knowledge.

Cognitive load theory is concerned with the acquisition of biologically secondary knowledge and is based on the cognitive architecture associated with the acquisition of such knowledge. Both human cognition and evolution by natural selection use an analogous architecture to process information (Sweller & Sweller, 2006). There are two ways humans can acquire novel information. We can either discover new, biologically secondary information during problem solving or we can obtain information from other people. Both these skills concerned with the acquisition of novel, secondary information are themselves biologically primary and do not need to be explicitly taught. We vastly exceed all other mammalian species in these two skills which explains, at least in part, the dominance of the human species in the mammalian world.

Following the acquisition of novel information, whether during problem solving or from others, it needs to be processed. It is processed by a severely limited capacity (Cowan, 2001; Miller, 1956) and duration (Peterson & Peterson, 1959) working memory. Once processed, information then can be stored in a long-term memory with no known limits of capacity or duration. As first demonstrated by De Groot (1965), expertise is domain specific and is due to enormous amounts of domain-specific information stored in long-term memory.

Lastly, signals from the environment can trigger the transfer of information from long term back to working memory to execute appropriate action. Unlike when dealing with novel information from the environment, there are no known limits to the amount of information that can be transferred back to working

memory or the amount of time required to hold and process that information. For example, if any reader of this paper sees the very complex text squiggles: “the black cat”, and is asked to reproduce them, they are likely to be able to do so flawlessly. Furthermore, if they are asked to reverse the order of the squiggles mentally to produce, “tac kcalb eht”, they are likely to also succeed in this task. Success in these tasks is entirely dependent on years of people using their cognitive architecture to learn to read and write English. Not only are there no known limits to the amount of information held in long-term memory, but there also are no known limits to the amount of stored and organised information that can be transferred back to working memory to allow a person to function in a complex environment. In this manner, education transforms us.

This human cognitive architecture provides a base for cognitive load theory. Because of our limited working memory when dealing with novel information, some categories of biologically secondary information can be readily processed by humans, whereas other categories can be very difficult to process. Cognitive load theory uses the concept of element interactivity to determine and describe these categories (Chen et al., 2023; Sweller, 2010). The elements of some types of information interact and so must be processed simultaneously in working memory due to high element interactivity, resulting in a high working memory load. When high element interactivity is caused by the natural complexity of information being processed, it is called intrinsic cognitive load. For example, when learning to solve an algebra problem, a change in any part of an algebraic expression is likely to require the entire expression to be considered, requiring attending to the elements’ interactivity and thus imposing a heavy working memory load. In contrast, when learning the vocabulary of a second language or learning the symbols of the chemical periodic table, each element can be learned in isolation without reference to any other element. Even if there are many elements, the task may be difficult, but due to its low element interactivity, it is not necessarily complex, thus resulting in a low intrinsic load that does not impose a high working memory load.

A second kind of load, extraneous cognitive load, refers to complexity imposed by how information is presented to learners and the cognitive activities required of them (Chen et al., 2023; Sweller, 2010). The effectiveness of instruction is influenced by extraneous cognitive load and in turn, extraneous cognitive load is determined by element interactivity. Additional extraneous load reduces the effectiveness of instruction and so instruction should be designed to minimise or eliminate that category of cognitive load. For example, as indicated below when discussing the worked example effect, studying worked examples demonstrating a solution to a problem results in superior knowledge for novice learners compared to attempting to solve the same problem oneself. The number of elements that need to be processed when attempting to solve a problem oneself is greater than the number of elements that need to be processed when studying an equivalent worked example.

Finally, germane cognitive load refers to working memory resources devoted to learning (Sweller, 2010). Because working memory resources that need to be devoted to learning are determined by intrinsic and extraneous cognitive load, no instructional consequences of germane cognitive load have been identified and so

germane cognitive load is no longer considered an independent source of load and the term is less commonly used.

The extraneous and intrinsic subcategories of cognitive load have been used to generate a variety of cognitive load effects, which have been tested using randomised, controlled trials. The history of these subcategories, and also germane cognitive load, is relevant to the theme of theory development. The distinction between extraneous and intrinsic cognitive load was made when empirical results indicated that cognitive load effects seemed to randomly appear and disappear (see the section “Failure to Replicate a Variety of Cognitive Load Theory Effects due to the Element Interactivity and Expertise Reversal Effects” below). In contrast to the development of extraneous and intrinsic cognitive load, the idea of germane cognitive load was not generated from data. It was developed because it seemed to be a plausible and interesting idea. Nevertheless, over the years it became clear that there were many extraneous and several intrinsic cognitive load effects, but there were no germane cognitive load effects being generated. With extraneous and intrinsic cognitive load able to explain all available data, some cognitive load theorists now refer to germane cognitive load as being the working memory resources used to handle intrinsic as opposed to extraneous cognitive load (Sweller, 2010). That change in definition means that germane cognitive load is no longer an independent source of cognitive load and so has a different status to extraneous and intrinsic load. This is an example of how theories develop and change over time (Greene, 2022).

Randomised, controlled trials have been used to validate cognitive load theory by generating hypotheses that were supported by data. The theory has generated a wide variety of instructional procedures, or instructional effects, based on these randomised, controlled trials. Some of those effects, along with the procedures used to generate them, are discussed below. Most of them were generated when earlier versions of the theory failed to replicate or explain the available data. Those earlier versions subsequently had to be modified as data relevant to the theory accumulated, culminating in the current version outlined above. The rest of this paper is concerned with some of the history of that process. It provides a case study of both replication failure and the integration of other theories, resulting in consequent theory development. I will begin with the origins of the current theory.

## Origins

In the early 1980s, my colleagues and I were running experiments on problem solving, best exemplified by a 1982 paper (Sweller et al., 1982). We gave our university student participants puzzle problems to solve in which they were presented a start number that they had to transform into a goal number, only using the two operations, multiply by 3 and subtract 69. They could use each operation as often as needed and in any sequence, but had to find a sequence that allowed them to attain the goal number. All arithmetic was carried out by a computer, so all the participants had to decide at each choice point was which of the two operations to carry out. All problems could be solved in between 2 and 10 moves. Ten-move problems were, of course, more difficult than 2-move problems but nevertheless, most participants had

little difficulty attaining solution on the several problems of multiple lengths that the participants were presented.

This ease of solution result did not surprise us because every problem could be solved only by an identical procedure of alternating the two allowed operations, starting by multiplying by 3 before subtracting 69, until the goal was attained. The number of moves required to reach the goal number changed but the solution rule remained constant because the initial and goal numbers had been chosen to ensure that no other solution was possible. Nevertheless, there was one surprise. When we tested to see whether participants had learned the alternation rule, very few of them seemed to be aware of it despite successfully solving multiple problems in which they had necessarily used the rule. Participants who were aware of the rule used it to solve problems that could be much more simply solved by alternative means thus demonstrating *Einstellung* (i.e. mental set effects; Luchins, 1942) by, for example, attempting to use the alternation rule to solve a simple problem that could be solved by multiplying by 3 twice in succession. In addition, participants who were aware of the rule could easily solve very complex problems requiring many moves by using the rule. Participants who were unaware of the rule did not demonstrate *Einstellung* and took longer to solve complex problems.

I cannot remember whether my colleagues were surprised by our results, but I was astonished. Participants were solving multiple problems following a rule of which they seemed to be oblivious. I became less surprised when, sometime after the paper was published, I introspected my own behaviour. As an avid puzzle problem solver, I remembered that I frequently spent long periods working on a puzzle, eventually arrived at a solution, but seemed unconscious of my procedure when I attempted to re-solve it.

Solving puzzle problems is rightly of little interest in education, but it struck me that this result might have educational implications. We did not inform our participants of the alternation rule because we wanted to study problem solving. Had we informed them of the rule, it would have taken a few seconds and they obviously could have effortlessly solved any length problem presented to them. Based on the results of these experiments, I realised if failure to learn a basic characteristic of the problems being solved is a possible consequence of problem solving compared to being explicitly told the rule, why use problem solving as a learning device in educational contexts? Why not reduce learning times by explicitly instructing students? Over 40 years later, I still seem to be periodically fighting this battle for direct instruction rather than asking students to engage in problem solving (Chen et al., 2023; Zhang et al., 2022). In the meantime, there were other issues to consider, the most important of which was what are the characteristics of human cognitive architecture that led to our experimental results?

## Some Aspects of Human Cognitive Architecture

The modal model of aspects of human cognitive architecture introduced by Atkinson and Shiffrin (1968) was well established by the 1980s. It included sensory, working, and long-term memories. From an educational perspective,

working and long-term memories are particularly critical and as can be seen above, remain central to cognitive load theory. Working memory is used to process information when engaging in activities such as problem solving. In contrast, long-term memory is used to store information for subsequent use. The relations between working and long-term memories could potentially explain the results of the Sweller et al. (1982) problem-solving experiments.

A transformation problem requires people to transform problem states into other states using problem-solving operators. The numerical transformation problems used by Sweller et al. (1982) provide an example. Working memory is used to find a transformation-problem solution using a procedure known as means-ends analysis (Newell & Simon, 1972) in which the problem solver attempts to find problem operations that will reduce the distance between the current problem state and the goal state. Having to simultaneously consider the current problem state, the goal state, and possible operators that will reduce the differences between the two states is very resource intensive (Sweller, 1988). Working memory is used in this problem-solving process, but working memory is very limited in capacity (Cowan, 2001; Miller, 1956) and duration (Peterson & Peterson, 1959). When limited working memory resources are used to move closer to the goal, those resources may be unavailable for other cognitive activities such as comparing the relation between multiple moves made previously during the solution process. That limited working memory may explain why problem solvers could readily reach the goal of a problem but have no idea how they had accomplished this aim.

The problem solvers in the Sweller et al. (1982) experiments may have been entirely concerned with deciding which moves to make without considering their relations to previous moves and so they remained oblivious to the alternation of the multiplication and subtraction operations being made to solve the problems. Accordingly, the problems could be solved with little or nothing being transferred to long-term memory for subsequent use. If so, there could be implications for the use of problem solving as a learning device in education.

The beginnings of cognitive load theory can be traced to the application of the cognitive architecture outlined by Atkinson and Shiffrin (1968) to the problem-solving results obtained by Sweller et al. (1982). The route was somewhat circuitous and as often happens, the milestones are not properly indicated in the literature and so rely on faulty memories. Nevertheless, by the time Sweller (1988) was published, the initial outline of cognitive load theory was established. The theory could be used to generate hypotheses to be tested using randomised, controlled trials. Successful, replicated experimental results became cognitive load effects that could be used in instructional design, but it became immediately apparent that successful replications were mixed with unsuccessful ones. In a very real sense, the unsuccessful replications were at least as important and possibly more important than the successful ones. The next sections indicate the variety of revisions that have been made to both basic cognitive load theory and to the instructional effects generated by the theory. Table 1 summarises those revisions.

**Table 1** Summary of revisions to cognitive load theory

Reason for the revision	Consequence of the revision
Failure to replicate the worked example effect using problems other than algebra transformation problems	Introduction of the split-attention effect
Failure to replicate the split-attention effect using sources of information intelligible in isolation	Introduction of the redundancy effect and the incorporation of the logical relation between sources of information into cognitive load theory
Failure to replicate the modality effect using lengthy spoken statements	Introduction of the transient information effect
Failure to replicate many cognitive load theory effects using low element interactivity information	Introduction of the expertise reversal effect
Inability to explain why explicit instruction was required for some tasks but not others	Incorporation of the distinction between biologically primary and secondary information into cognitive load theory
Elucidation of an analogy between the information processing characteristics of evolution by natural selection and human cognition	Suggestion that both evolution by natural selection and human cognition are natural information processing systems, providing a context for cognitive load theory

## Failure to Replicate the Worked Example Effect

The worked example effect was one of the earliest cognitive load effects generated. If, as is still central to cognitive load theory, problem solving can overwhelm working memory resulting in less information being transferred to long-term memory, then learning to solve algebra problems such as  $(a + b)/c = d$ , solve for  $a$ , might be easier by studying worked examples demonstrating the solution rather than solving the problems. The number of elements that need to be processed by working memory when studying a worked example should be reduced compared to solving the equivalent problem on one's own. In the terminology used by the current version of the theory, solving a problem imposes an extraneous cognitive load. Using algebra problems, Sweller and Cooper (1985) and Cooper and Sweller (1987) provided evidence from multiple experiments supporting the hypothesis with successful replications. Demonstrating that studying worked examples resulted in higher test scores than solving the equivalent problems appeared to be a very robust finding.

The obvious next step was to confirm the experimental results using different problems. We tried geometry problems and obtained no evidence of superiority when students studied worked examples instead of solving problems. We tried kinematics problems in physics classes and again found no worked example effect. In those days, replication failure was not an issue that was commented on in the literature, but we had our own, very strong examples of multiple, replication failures. These failures persisted for several years and appeared to signal the early demise of cognitive load theory in general and the worked example effect in particular.



## The Split-Attention Effect

Our worked example replication failure seemed not to be due to any specific experimental design or procedural flaw. We needed to go back to the theory. Why could our algebra worked examples reduce working memory load compared to solving the equivalent problems, but geometry or kinematics worked examples not have the same effect? Ironically, the answer could be found in the very basis of early versions of cognitive load theory. We had assumed that any worked example, irrespective of its structure, would reduce cognitive load. Of course, at one extreme, it is always possible to design a worked example or any instructional information that is unintelligible to most learners. We had not gone to that extreme, but it turns out that the conventional structure of algebra worked examples accords with basic cognitive load theory principles. The conventional structure used for geometry and physics worked examples does not.

To understand an algebra worked example, learners must attend to each line and ensure they understand the algebra that transforms that line to the next line. To understand the solution to  $a + b = c$ , *solve for a*, learners need to understand the move from  $a + b = c$ , to  $a + b - b = c - b$ . The 2 lines are usually presented using the following format:

$$\begin{aligned} a + b &= c \\ a + b - b &= c - b \end{aligned}$$

Understanding that move requires attending to those two lines and nothing else. The normal structure of such worked examples does not result in an unnecessary increase in elements of the information that need to be processed. In other words, it does not impose an extraneous cognitive load.

In contrast, consider a conventional geometry worked example. Commonly, it will consist of a geometric diagram along with statements beneath the diagram in the form of *angle ABC = angle XYZ (relevant theorem)*. In isolation, the diagram alone does not reveal the problem solution unless one has already learned to solve such problems. The written statements that need to be associated with the diagram are unintelligible in isolation and only can be rendered intelligible by referring to the diagram. Each statement must be mentally integrated with the diagram. Consequently, working memory must simultaneously process and relate the information contained in both the diagram and the statements. Each time a statement such as *angle ABC* is read, the reader must switch attention from the statements to search for *angle ABC* in the diagram. Once the angle is found, attention must be switched back to the list of statements. The extraneous working memory load may be insurmountable due to learners having to split their attention between the diagram and the statements and then mentally integrate them to render the worked example intelligible.

To reduce the working memory load of geometry worked examples, the diagram and the statements need to be physically integrated either by placing statements such as *angle ABC* in or next to the actual angle of the diagram, or by having an arrow lead from the statement to the actual angle. In that way, learners do

not have to search for referents in the diagram, thus reducing extraneous cognitive load. Data indicated that the worked example effect could readily be obtained provided geometry and kinematics worked examples were structured to reduce working memory load by reducing or eliminating split attention. A suitable restructure consisted of physically integrating the multiple sources of information to reduce or eliminate the need for learners to search for referents. Examples of split-attention and integrated worked examples may be found in Tarnizi and Sweller (1988) and Ward and Sweller (1990). Physically integrating worked examples reduced split attention and reinstated the worked example effect by reducing working memory load, thus expanding cognitive load theory.

### **Failure to Replicate the Split-Attention Effect**

The split-attention effect indicated that in general, if learners had to split their attention between multiple sources of information and mentally integrate them, and specifically, if those multiple sources of information consisted of diagrams and text, the need to split attention and mentally integrate would increase extraneous cognitive load and reduce learning. Physical integration would eliminate the need for mental integration and so reduce cognitive load and facilitate learning. It was a short and obvious next step to assume that all diagrams and related text needed to be physically integrated.

We were disabused of that notion by our failure to replicate the split-attention effect using different types of information. Multiple experiments indicated that physically integrating some diagrams and text yielded no learning gains. Just as we had failed to replicate the worked example effect under split-attention conditions, now we were failing to replicate the split-attention effect using some combinations of diagrams and text but succeeding in demonstrating the effect using other diagrams and text. We had no idea why we were failing to replicate the split-attention effect. Eventually, we realised it was due to the redundancy effect.

### **The Redundancy Effect**

Consider again a geometry diagram along with a set of statements indicating a solution to a problem based on that diagram. The logical relations between the diagram and the statements are critical. For a learner new to the area, the information in neither source of information can function as a usable worked example in the absence of the other source of information. References to “angle ABC” in a statement only become meaningful in conjunction with the relevant diagram. To be intelligible as a worked example, the two sources of information must be integrated either mentally or physically. It can be easily assumed that any diagram with related verbal information has similar properties. That assumption is erroneous.

Instead of a diagram and text that are uninformative in isolation, consider a diagram that provides all the information that is required to understand and learn along with textual information providing the same information that is equally intelligible

in isolation as the diagram. An example is a diagram of the flow of blood in the heart, lungs, and body in which the direction of flow is indicated by arrows that, for example, indicate that blood flows from the right ventricle to the pulmonary artery. Statements associated with the diagram can be of the form “Blood flows from the right ventricle to the pulmonary artery.” These statements can be either beneath the diagram using a split-attention format or integrated with the diagram using an integrated format. Chandler and Sweller (1991) conducted multiple experiments with a variety of materials including the blood-flow information and did not find evidence for the split-attention effect. Physically integrating the diagrams and the text rather than having the text under the diagrams to supposedly reduce split-attention did not provide any advantage. Instead, Chandler and Sweller (1991) obtained evidence that learning was facilitated by the elimination of the written information rather than by its integration with the diagrams. This failure to find evidence of improved learning following the integration of diagrams and text provides another example of conceptual replication failure.

Chandler and Sweller (1991) had difficulty replicating the split-attention effect, but they found a new effect, the redundancy effect, that allowed another expansion of cognitive load theory. The exact relation between different sources of information is critical to instructional design. Simply observing the surface structure of information is insufficient. The logical relations between sources of information must also be considered. Physically integrating two sources of information that are unintelligible in isolation reduces working memory load and facilitates learning due to the reduction of split attention. Physically integrating two sources of information where one source is unnecessary to understanding increases rather than decreases working memory load due to the extraneous cognitive load imposed by redundancy. Eliminating such unnecessary information reduces working memory load and facilitates learning due to the elimination of redundant information. Our failure to replicate the split-attention effect gave birth to the redundancy effect and an expansion of cognitive load theory.

## Failure to Replicate the Modality Effect

The modality effect occurs when pairing a visual source of information such as a diagram with a spoken and hence auditory source of information enhances learning compared to presenting the spoken information in written form instead (Mousavi et al., 1995). From the perspective of human cognitive architecture, the modality effect is assumed to occur because visual and auditory processors are partially independent. If all information is presented in visual form as occurs when a diagram is paired with written text, the visual processor may be overloaded. Transferring some of the information to the auditory processor by using spoken rather than written text may reduce the load on the visual processor and so enhance learning.

The same rules that apply to the split-attention effect concerning the logical relations between the two sources of information also apply to the modality effect. The two sources of information must be unintelligible in isolation and so must be considered in conjunction. If the verbal information merely replicates diagrammatic

information, or if it is entirely unrelated to the diagrammatic information, it is redundant and so redundancy effect rather than modality effect conditions apply. We had learned from our findings concerning the redundancy effect that adding, for example, redundant text was not going to be beneficial irrespective whether that text was presented in written or spoken form. Under redundancy conditions, performance will be enhanced by the elimination of one source of unnecessary information rather than by converting written into spoken text. If changing the modality of text presentation was going to be effective, it presumably had to be under split attention rather than redundancy conditions. Nevertheless, it turns out that a blanket rule to use dual-modality instruction under split-attention conditions does not work either, leading to the transient information effect.

### The Transient Information Effect

There is considerable evidence for the modality effect (Ginns, 2005). Nevertheless, there were published failures that were particularly striking because they did not just fail to obtain the modality effect, they obtained a reverse modality effect in which the single modality group that was presented all of the information in visual form obtained higher test scores than the dual modality group (Tabbers et al., 2004). There was no obvious, satisfactory explanation for this replication failure. Nevertheless, it was not an isolated result. In the first of two experiments, Leahy and Sweller (2011) replicated this failure to obtain the modality effect. It was not due to redundancy because neither Tabbers et al. nor Leahy and Sweller had used materials that incorporated redundant information.

Leahy and Sweller (2011) argued that the reason some experiments reversed the modality effect was because of the length of the auditory component. Auditory information, unlike written information, is transient. When listening to speech, any information heard now immediately disappears to be replaced by new information. Because of this continual replacement of previous information by current information, the only way speech can be processed is by retaining previous information in working memory while simultaneously processing current information. For complex, lengthy text, the working memory load can easily exceed available capacity. In other words, speech can impose a heavy cognitive load and indeed, beyond a certain point, an impossible cognitive load. In contrast, written information is permanent which is why, of course, writing was invented. People can read text, and because it is permanent, they can return to it repeatedly. Speech cannot be readily returned to, although it is becoming easier using modern technology.

Leahy and Sweller (2011) hypothesised that the reverse modality effect occurred because complex, lengthy speech may overwhelm working memory. The same information presented in written form may be much more readily processed. Accordingly, a reverse modality effect should sometimes be obtained with written text plus diagrams being superior to spoken text plus diagrams.

After obtaining a reverse modality effect in which visual-only information was superior to dual modality information, in a second experiment, Leahy and Sweller (2011) used much shorter text and obtained a conventional modality effect with

dual auditory and visual modality information superior to information presented in written form only. Once again, a failure to replicate experimental results could be accounted for by an expansion of the relevant theory. Lengthy, complex text in spoken form can be difficult or impossible to process. The same information presented in written form can reduce cognitive load, providing an explanation for the reverse modality effect.

## **Failure to Replicate a Variety of Cognitive Load Theory Effects due to the Element Interactivity and Expertise Reversal Effects**

The concept of element interactivity was developed within cognitive load theory to explain why some information can be difficult to process and learn (Chen et al., 2023; Sweller, 1994, 2010). It was not part of the initial versions of the theory but now is a central concept of the theory, determining working memory load due to the intrinsic nature of the information being processed (i.e., intrinsic cognitive load) or the nature of the instructional procedures used (i.e., extraneous cognitive load). As indicated in the current version of the theory above, some information consists of elements that closely interact in the sense that learning cannot proceed without simultaneously processing all the elements, resulting in high element interactivity. In contrast, other information consists of multiple elements that can be processed in isolation without referring to each other, resulting in low element interactivity.

Element interactivity cannot be determined just by reference to the structure of the task. The expertise of learners also determines element interactivity. The structures and functions of human cognitive architecture alter what constitutes an element depending on knowledge held in long-term memory. For example, while an algebra equation and problem such as  $(a + b)/c = d$ , solve for  $a$  consist of multiple, interacting elements for students learning algebra, for most readers of this paper, that problem constitutes a single element. Some people are highly familiar with problems of that type and fully understand the problem and its solution. They can transfer appropriate knowledge concerning the problem from long-term to working memory as a single element. That function of human cognitive architecture (i.e., chunking information) has instructional implications.

Summarising at this point, element interactivity can be varied in two ways. First, some information is low in element interactivity because the various elements can be processed without reference to each other. In contrast, other information is high in element interactivity because it consists of many interacting elements. Second, element interactivity can be varied by increasing expertise with respect to particular information (Chen et al., 2023; Sweller, 2010). If knowledge of interacting elements is held in long-term memory, they can be treated as a single element resulting in high element interactivity information being turned into low element interactivity information.

The change in element interactivity with expertise has led to cognitive load effects. Most of those effects assume learners are novices with limited knowledge of what is being taught. As expertise increases, multiple elements are combined into a single element and so the cognitive load effects reduce in size, resulting in what

is known as the element interactivity effect. Because of the element interactivity effect, with increasing expertise, cognitive load effects may disappear and may even reverse. This reversal is itself a cognitive load effect called the expertise reversal effect (Kalyuga et al., 2003), which is a specific example of the element interactivity effect (Chen et al., 2017). The expertise reversal effect occurs when increased levels of expertise reduce element interactivity and so an effect that occurs under high element interactivity conditions for novices disappears under low element interactivity conditions when expertise is increased.

Element interactivity as a measure of complexity cannot be accurately measured because it relies on knowledge of the exact contents of long-term memory, for any individual learner, but it can be estimated (Chen et al., 2023). The element interactivity and expertise reversal effects apply to all other cognitive load theory effects and so provide many examples of a failure to replicate. Those failures resulted in a theoretical advance. As an example, the worked example effect discussed above only is obtained for high element interactivity information. In accordance with the expertise reversal effect, a reverse worked example effect with problem solving being superior to studying worked examples is obtained for low element interactivity information (Chen et al., 2015, 2016a, b, 2017; Kalyuga et al., 2001). If levels of expertise are not considered, exactly the same instructional procedures can result in vastly different outcomes, but these different outcomes only give the impression of replication failure. In fact, of course, these are not replication failures, they are theory failures due to not recognising the critical variable of element interactivity. That variable is recognised in the current version of the theory and so the replication failure disappears.

## Developing Cognitive Load Theory via Theory Integration

Replication failure has provided a major impetus for the development of cognitive load theory, but there have been other sources of development. Initially, the theory was based on our knowledge of human cognitive architecture, specifically our knowledge of working memory, long-term memory, and the relations between them. More recently, evolutionary psychology has provided a rich source of cognitive load theory development, resulting in the current version of the theory outlined above. The relatively recent influence of evolutionary psychology on cognitive load theory has resulted in many refinements to the latter.

## Geary's Evolutionary Educational Psychology

As indicated above, cognitive load theory was designed to deal with instructional issues and only some categories of information are amenable to instruction. Geary's (Geary, 2002, 2005, 2008, 2012; Geary & Berch, 2016) categorisation of knowledge into biologically primary and secondary knowledge allowed the elimination of primary knowledge from the categories of knowledge that required instruction. From an instructional perspective, the importance of Geary's distinction between

biologically primary and secondary knowledge cannot be overestimated. For many years, there has been an emphasis on providing students with minimal guidance (Kirschner et al., 2006) and having them engage in discovery, problem-based, or inquiry learning. The immediate impetus for this movement came from Bruner (1961) although the general idea can be traced back to Dewey (1938) and even Rousseau (1762/1979). It was argued that the natural way to learn was to discover concepts and procedures for oneself rather than having them presented during explicit instruction. In other words, these theorists were implicitly assuming that engaging in species-typical childhood behaviours would promote biologically secondary academic learning (e.g., reading) in the same way they contribute to the unfolding of primary abilities (e.g., spoken language). It was a strong, interesting, and plausible argument that swept all before it.

I found the argument just as convincing as most other people, but I had a problem. When cognitive load theory-based researchers ran experiments on the worked example effect using school-based problems, having novice learners study worked examples frequently proved superior to actually solving the equivalent problems (Cooper & Sweller, 1987; Paas, 1992; Paas & van Merriënboer, 1994; Sweller & Cooper, 1985). If discovering concepts and procedures enhanced learning more than explicit instruction, why did randomised, controlled trials indicate the opposite result? Despite not exactly being a replication failure—those in favour of discovery learning seemed to have an aversion to running randomised, controlled trials—the worked example effect does provide the reverse result to that expected by problem-based learning.

Geary's evolutionary theorising provided the missing piece of the jig-saw puzzle. Schools were teaching an entirely different category of information from much that is learned external to formal education and that was why the information needed to be explicitly taught. Students cannot process and learn biologically secondary information in the same manner that they learn biologically primary information. If both categories of information were learned in the same way, societies would not have needed to establish educational institutions. The incorporation of Geary's distinction between biologically primary and secondary knowledge has considerably strengthened the cognitive architecture that underlies cognitive load theory.

### **The Analogy Between the Information Processing Characteristics of Evolution by Natural Selection and Human Cognition**

The information processing characteristics of human cognitive architecture and evolution by natural selection when acquiring biologically secondary knowledge are analogous (Sweller, 2022; Sweller & Sweller, 2006). Both provide examples of natural information-processing systems. Evolution by natural selection is normally considered a biological theory rather than a natural information processing system. It is both. Indeed, in these days of artificial intelligence, biological evolution can also be considered an intelligent system with capabilities that far exceed either human or artificial intelligence. The consequence is a creative system that vastly exceeds the creativity of either humans or artificial intelligence systems.

As described in the current version of cognitive load theory, there are two ways for humans to acquire information: either during problem solving using a random generation and test process which is analogous to random mutation in biological evolution, or from other people, equivalent to reproduction in biological evolution. Once acquired, information must be processed by a limited capacity, limited duration working memory, analogous to the epigenetic system that, among other functions, can facilitate or depress mutations in different locations. After being processed in working memory, information can be stored in a long-term memory with no known limitations just as unlimited amounts of stored information can be stored in a genetic code. The human cognitive system can retrieve information from long-term memory to govern action that is appropriate to the extant environment in a manner analogous to the way in which the epigenetic system uses environmental cues to turn genes on or off.

Cognitive load theory uses this cognitive architecture as its base to generate its instructional procedures. The fact that this base reflects what appears to be a natural information-processing system shared by evolution by natural selection strengthens the theory. Evolution is a creative engine that far exceeds anything that either human or artificial intelligence has accomplished. It has created every species and every complex process of life on earth. That creativity has been accomplished by a combination of a large store of readily transmissible information initially built by a random generate and test process but then transmitted indefinitely to descendants. The same basic procedures are central to human cognition. For humans to discover novel information is a long, slow journey of random generate and test during problem solving, but once discovered, humans have evolved to efficiently transmit the information to others.

The instructional goal of education should be building stores of relevant information in long-term memory for later use. It is that knowledge that permits critical and creative thinking, whereas attempting to teach these generic-cognitive skills has proven unsuccessful (Tricot & Sweller, 2014). Cognitive load theory emphasises that knowledge building should be the major aim of instruction. The analogy between the information-processing procedures of human cognitive architecture and evolution by natural selection underpins that emphasis.

## **Virtues and Contributions of Cognitive Load Theory, Its Current Status, and Future Directions**

All theories have boundary conditions (i.e. the scope and limits of a theory's applicability; Scheel et al., 2021), with the presence of new boundary conditions beyond the horizon always a possibility. We have used the exploration of boundary conditions to expand cognitive load theory. Several such examples of cognitive load theory expansion have been provided in this paper. I believe this process provides a positive example of theory development rather than a negative example of replication failure caused by faulty experimental procedures. Cognitive load theory's continual adaptation to new data is one of its primary virtues. The fact that those



data are collected primarily from full experiments using randomised, controlled trials with real students studying materials from their own curricula is also a virtue.

Empirical findings are and should be a major source of theory development, but they are not the only source. Cognitive load theory has continually incorporated other theories that accorded with its basic principles, providing another virtue. The theory began by relying heavily on what is known of human cognitive architecture (Sweller, 1988) by emphasising relations between working and long-term memory and followed by incorporating Geary's views (Geary, 2012) on categorising information into biologically primary information that humans have evolved to acquire and secondary categories that are central to educational science. In turn, the theory now additionally suggests that the information processing characteristics of the human cognitive system when dealing with biologically secondary information are analogically related to the information-processing characteristics of evolution by natural selection (Sweller, 2022). If this analogy is valid, the factors that are relevant to instructional design can be refined.

Expansion of cognitive load theory by the resolution of apparently contradictory findings and by the incorporation of additional concepts, procedures, and distinctions is an ongoing, continuous procedure that I feel should be a permanent aspect of all theories in educational psychology. A recent example of that process can be seen in the special issue of the *British Journal of Educational Psychology* on integrating other theories with cognitive load theory—see Hanham et al. (2023) for the editorial introduction. The papers in that issue cover level of expertise, cognitive load measurement, embodied cognition, self-regulated learning, emotion induction, replenishment of working memory, and subprocesses of working memory. All are concerned with promising variations to cognitive load theory based on the results of randomised, controlled trials. With the passage of time allowing additional data and theorising, such continuing theory development can be expected, with positive consequences for both the theory and instructional procedures.

A major contribution of cognitive load theory has been to alert the field of instructional design to some very basic and very well-known characteristics of human cognition. It has been known since at least Miller (1956) that working memory has severe limits and at least since De Groot (1965, first published in 1946) that skilled performance in complex areas required the storage of immense amounts of domain-specific information in long-term memory. In the past, that knowledge has had limited effects on instructional design principles with many common recommendations proceeding as though the characteristics and intricate relations between working memory and long-term memory do not exist, or if they do exist, are irrelevant to instruction (Kirschner et al., 2006; Zhang et al., 2022).

That lack of knowledge concerning basic human cognitive architecture is rapidly dissipating and consequently, the theory's acceptance and visibility are rapidly increasing. Based on Web of Science and Scopus databases, there are several thousand academic papers that have used the theory and many millions of references to cognitive load theory using Google search. Several major educational jurisdictions recommend the theory (e.g. New South Wales Department of Education 2017; Victorian Department of Education 2020; Perry et al. 2021). In addition, there now are books for teachers on cognitive load theory (Garnett, 2020; Lovell, 2020).

Current work is being carried out on working memory depletion after cognitive effort and recovery after rest (Chen et al., 2021). An implicit assumption of cognitive load theory has been that working memory capacity is relatively stable, other than variations caused by knowledge held in long-term memory. As has occurred repeatedly with other assumptions of the theory, that assumption may only be partially correct. For example, some versions of the spacing effect, which occurs when information processed with rest periods facilitates learning compared with the same information presented in massed form, may be due to working memory depletion after cognitive effort and recovery after rest (Chen et al., 2018, 2021). Spacing may provide rest periods allowing working memory recovery that is not available under massed learning conditions. Further work on this “resting effect” will need to be carried out in the future to verify the effect, establish its limits, and then inform further cognitive load theory development.

## Conclusion

In conclusion, cognitive load theory has been under constant development for over 40 years. It has developed via epistemic iteration after replication failure and theory integration. The need for permanent theoretical development caused by the enormous number of variables with which researchers must deal is unlikely to just apply to cognitive load theory and the issues with which the theory is concerned. It may be characteristic of the field of educational psychology.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions

**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

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). Academic Press.
- Bruner, J. (1961). The art of discovery. *Harvard Educational Review*, 31, 21–32.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8, 293–332.
- Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, 107, 689–704.

- Chen, O., Kalyuga, S., & Sweller, J. (2016a). Relations between the worked example and generation effects on immediate and delayed tests. *Learning and Instruction, 45*, 20–30.
- Chen, O., Kalyuga, S., & Sweller, J. (2016b). When instructional guidance is needed. *Educational and Developmental Psychologist, 33*, 149–162.
- Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review, 29*, 393–405. <https://doi.org/10.1007/s10648-016-9359-1>
- Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. *Educational Psychology Review, 30*, 483–501. <https://doi.org/10.1007/s10648-017-9426-2>
- Chen, O., Paas, F., & Sweller, J. (2021). Spacing and interleaving effects require distinct theoretical bases: A systematic review testing the cognitive load and discriminative-contrast hypotheses. *Educational Psychology Review, 33*, 1499–1522.
- Chen, O., Paas, F. & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review, 35*. <https://doi.org/10.1007/s10648-023-09782-w>
- Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*, 347–362.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–114.
- De Groot, A. (1965). *Thought and choice in chess*. Mouton. (Original work published 1946).
- Dewey, J. (1938). *Experience and education*. Macmillan.
- Garnett, S. (2020). *Cognitive load theory: A handbook for teachers*. Crown House Publishing.
- Geary, D. (2002). Principles of evolutionary educational psychology. *Learning and Individual Differences, 12*, 317–345.
- Geary, D. (2005). *The origin of mind: Evolution of brain, cognition, and general intelligence*. American Psychological Association.
- Geary, D. (2008). An evolutionarily informed education science. *Educational Psychologist, 43*, 179–195.
- Geary, D. (2012). Evolutionary educational psychology. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 1, pp. 597–621). American Psychological Association.
- Geary, D., & Berch, D. (2016). Evolution and children's cognitive and academic development. In D. Geary & D. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp. 217–249). Springer.
- Gehlbach, H., & Robinson, C. D. (2018). Mitigating illusory results through preregistration in education. *Journal of Research on Educational Effectiveness, 11*, 296–315.
- Gehlbach, H., & Robinson, C. D. (2021). From old school to open science: The implications of new research norms for educational psychology and beyond. *Educational Psychologist, 56*, 79–89.
- Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313–331.
- Greene, J. A. (2022). What can educational psychology learn from, and contribute to, theory development scholarship? *Educational Psychology Review, 34*, 3011–3035.
- Hanham, J., Castro-Alonso, J., & Chen, O. (2023). Integrating cognitive load theory with other theories, within and beyond educational psychology. *British Journal of Educational Psychology, 93*(Supplement 2), 239–250.
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Irvine, E. (2021). The role of replication studies in theory building. *Perspectives on Psychological Science, 16*, 844–853.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579–588.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23–31.
- Kirschner, P., Sweller, J., & Clark, R. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.
- Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*, 943–951.
- Lovell, O. (2020). *Sweller's cognitive load theory in action*. John Catt.

- Luchins, A. (1942). Mechanisation in problem solving: The effect of Einstellung. *Psychological Monographs*, 54(Whole No. 248).
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319–334.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice Hall.
- Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26, 1596–1618.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84, 429–434.
- Paas, F., & van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86, 122–133.
- Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193–198.
- Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, 56, 90–100.
- Rousseau, J.-J. (1762/1979). *Emile: or On Education*. Basic Books.
- Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, 16, 744–755.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4, 295–312.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, 22, 123–138.
- Sweller, J. (2022). The role of evolutionary psychology in our understanding of human cognition: Consequences for cognitive load theory and instructional procedures. *Educational Psychology Review*, 34, 2229–2241. <https://doi.org/10.1007/s10648-021-09647-0>
- Sweller, J., & Cooper, G. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition & Instruction*, 2, 59–89.
- Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, 4, 434–458.
- Sweller, J., Mawer, R. F., & Howe, W. (1982). Consequences of history-cued and means-end strategies in problem solving. *American Journal of Psychology*, 95, 455–483.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
- Sweller, J., van Merriënboer, J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31, 261–292.
- Tabbers, H., Martens, R., & Van Merriënboer, J. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, 74, 71–81.
- Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, 80, 424–436.
- Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review*, 26, 265–283. <https://doi.org/10.1007/s10648-013-9243-1>
- Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, 7, 1–39.
- Zhang, L., Kirschner, P., Cobern, W., & Sweller, J. (2022). There is an evidence crisis in science education policy. *Educational Psychology Review*, 34, 1157–1176.

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