Ancillary diagrams: A substitute for text in multimedia resources?

Richard Lowe[1,2] and Jean-Michel Boucheix[1]

[1] University of Burgundy Franche-Comte (LEAD-CNRS), Dijon, France
[2] Curtin University, Perth, Australia

Abstract.
Multimedia resources conventionally combine descriptive and depictive representations to convey their subject matter. However, responsibility for explaining that content is typically skewed heavily towards multimedia's descriptive components. This theoretical paper considers likely perceptual and cognitive processing requirements for internalizing these two sources of information during mental model construction. It uses the example of a multimedia resource consisting of written text and an accompanying overview picture to propose that much of the role currently allocated to text in such a resource could conceivably be reallocated to a set of ancillary diagrams. This proposal is based on an analysis suggesting that these diagrams are a better foundation for mental model building than is text. As a consequence, replacing the text in a multimedia resource with appropriately designed ancillary diagrams should result in superior understanding. Likely benefits and costs of this approach as well as possibilities for its further development are discussed.

Keywords: Multimedia explanation, Internal and external representations, Ancillary diagrams, Mental model construction.

1 Introduction

Our current information-rich society is one in which visual forms of information are increasingly pervasive. Along with this rise in our reliance on such visualizations, the way that society uses verbal information has also been changing significantly. These changes are particularly evident in the growth of short form text-based communications (mobile phone text messaging, social media platforms, news websites, etc.). There has also been a complementary rapid uptake of various types of static and dynamic visuals across these avenues of communication. However, one area that has lagged somewhat behind this trend is that of explanatory multimedia (as commonly found with technology-based educational and training resources, product manuals etc.). In multimedia resources, words (written or spoken) still typically carry the primary responsibility for presenting information to the target audience. Although it is certainly true that such resources are usually generously illustrated these days, the included pictorial material is rarely relied on for conveying the bulk of the content – that remains largely the job of the text.

In this chapter, we propose that the 'over-reliance' on text-based explanation evident in much current multimedia design may limit a resource's effectiveness for developing understanding. Historically, researchers and practitioners have advocated replacing text with pictures for audiences who are not native speakers of a local language or whose language skills are otherwise deficient [1]. However, this is not the theme of the present chapter. We do not propose doing away with text because the target audience lacks a basic level of comprehension of text *per se*. Instead, we argue for visualizations to be given greater explanatory responsibility than they have had hitherto, and for the role of text used together with explanatory visuals to be reconceptualized. Our proposal is that static or animated diagrams (provided that they are properly designed) could do much the same job as has traditionally been done by explanatory text, and that the purpose of text could be changed to one of supporting visual interpretation of these ancillary diagrams. The rationale for suggesting such changes is based on differences in the extent to which the fundamental characteristics of descriptive (textual) and depictive (pictorial) representations align with those posited for mental models [2,3]. We raise the possibility that our proposed changes to the design of multimedia resources could substantially improve their effectiveness for fostering understanding.

1.1 An example

The example of a manual caulking gun (Figure 1) will be used to illustrate the proposal being put forward in this chapter.

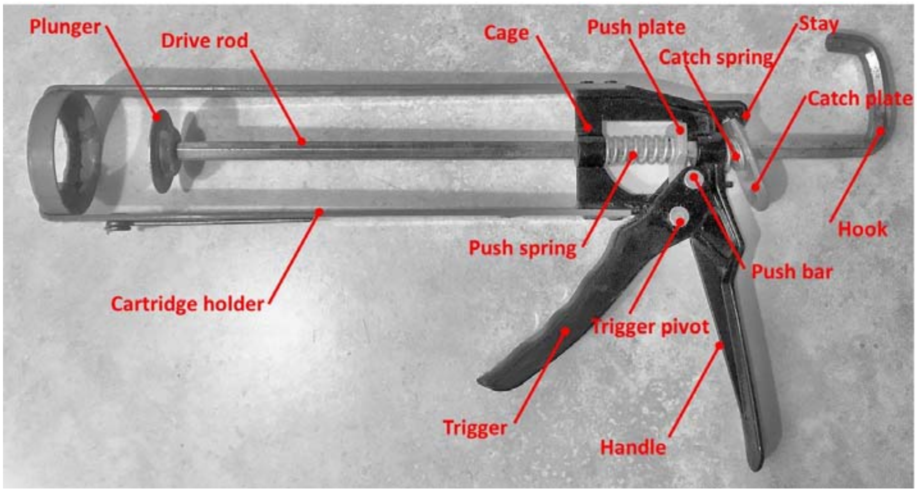


Fig. 1. Caulking gun overview picture

Let us suppose for the purposes of our discussion that we wish to develop a multimedia resource to explain how the mechanism of a caulking gun allows this device to perform its overall function. Caulking guns are typically used in housing construction and maintenance to extrude a continuous bead of viscous caulking compound (such as silicone sealant) that fills gaps between adjacent non-mating surfaces. After a cartridge of caulking compound is fitted into the gun, its contents are progressively extruded by means of successive squeezes of the trigger. Successful functioning of the caulking gun system relies on the interaction of the mechanism's two main sub systems:

1. The 'push' sub system that moves the drive rod through the cartridge so that its plunger pushes out the caulking compound. With each squeeze of the trigger, this sub system ejects a dose of the cartridge contents.
2. The 'catch' sub system that ensures the drive rod progressively moves through the cartridge so that all the contents are ultimately ejected. It does this because the grip that the catch plate exerts on the drive rod prevents the rod from slipping backwards after each trigger press.

The 'push' and 'catch' sub systems perform their individual roles that together contribute to the mechanism's proper overall functioning via two complementary causal chains. Each of these chains consists of a series of components that propagate activity between primary cause and ultimate effect by means of inter-component contact interactions. For example, consider the chain of events that occurs when the initial *cause* (a squeeze of the trigger) leads to the final overall *effect* (extrusion of some caulking compound). When the trigger is given its first squeeze, the 'push' causal chain begins with depression of the trigger towards the fixed handle of the gun. As the pivoted trigger rotates, the movement of its push bar in contact with the push plate causes that plate to change from a vertical to an angled orientation. This angling of the push plate in turn causes it to grip the drive rod then push it a limited distance along inside the cylinder so that the first dose of the caulking compound is ejected by the plunger. In concert with the operation of this 'push' causal chain, a second parallel 'catch' causal chain operates by which the tendency of the drive rod to retreat to the original position it occupied before the trigger squeeze (*cause*) is counteracted by grip from the catch plate that arrests its movement (*effect*). Successful functioning of the caulking gun mechanism depends on a coordinated and finely calibrated interplay between these two causal chains. Fundamental to this interplay is the relationship of a long, strong push spring (impinging on the push plate) to a short, weaker catch spring (impinging on the catch plate).

It is clear from the above account that despite the caulking gun being a common, easily operated device, the mechanism responsible for the gun's functionality is rather sophisticated. From the point of view of comprehending precisely how this mechanism works, this actually makes it quite complex. Consequently, designing an effective explanation to help people fully understand the way its numerous individual components (about a dozen of them) contribute individually and collectively to its overall functioning presents a considerable challenge. Currently, a popular response to such a challenge would be to develop a multimedia resource that uses a combination of descriptive and depictive information to explain this content.

This 'text plus picture' approach to multimedia has several common present day variants that can involve modifications such as using spoken rather than written text or using animated rather than static pictures [4]. Nevertheless, the way that multimedia resources of this type are currently designed still typically reflects their heritage from traditional printed textbooks where text was the primary carrier of information and pictures were generally treated as subservient adjuncts to the text=based explanation. In the next section, we consider various types of representation on the basis of a relatively straightforward multimedia implementation consisting of a written text that explains the functioning of a caulking gun accompanied by an overview picture (as per Figure 1) that shows its main parts.

2 Representations

Representations have a 'stand-for' relationship with their referents (i.e., the subject matter to which they refer). They can be either external to a person (such as printed and spoken text, or static and dynamic pictures) or internal (such as mental images, propositional knowledge, or mental models) [5,6,7]. Comprehension of the subject matter referred to by external representations requires the operation of 'bottom-up' perceptual and cognitive processes to extract then internalize relevant aspects of the available information. This internalized information is complemented 'top-down' by stored knowledge gained from prior experience to construct a mental representation of the referent subject matter. Ideally, this mental representation is a coherent knowledge structure that captures the subject matter sufficiently well to act as a basis for successful task performance.

Table 1 summarizes some key attributes of the types of representation that are the focus of this paper: mental models, depictive representations (pictures), and descriptive representations (text). It should be referred to when reading the following sections.

	Mental Model (internal tokens/relations)	Depictive (external picture)	Descriptive (external text)
Structure	Non-linear	Non-linear	Linear
Interrogation	Not prescribed	Not prescribed	Conventional reading direction (L to R; T to B)
Information Type	'Quasi-visuospatial'	Visuospatial	Non visuospatial
Representation	Analog but tokenized	Analog	Propositional
Processing Route	Mental construction	Direct processing as visuospatial	Multi-step: Interpretation then conversion into analog

Table 1. Comparison of mental models, depictive and descriptive representations

2.1 External versus Internal Representations

Mental models are internal knowledge structures that individuals construct via bottom-up and top-down processing. A good understanding of a topic is assumed to be the result of constructing a high quality mental model. These internal representations have been posited to consist of mental tokens and relationships that are organized in an analog, quasi-visuospatial manner that reflects the information structure of the subject matter they represent [6,7]. A high quality mental model is one having a close correspondence with its referent subject matter and that therefore makes it very useful in tasks such as prediction and inference. One crucial factor that can determine the quality of a mental model formed from an external representation is the level of processing challenge involved in internalizing that information source [8,9,10]. More specifically, mental model construction is likely to be facilitated if differences between characteristics of the type of external representation upon which it will be based and the characteristics posited for mental models are kept to a minimum. The greater the differences between these two classes of representation, the more demanding will be the processing required. In particular, if an external representation requires a substantial amount of preliminary processing in order to make the information it is carrying readily compatible with the requirements for mental model building, the likely outcome will be a lower quality internal representation.

2.2 Descriptive versus Depictive Representations

The words comprising a piece of text are constituted from agreed sub sets of alphabetic symbols organized in a particular sequence. Groupings of these words are in turn are arranged according to rules of syntax and semantic constraints. Their physical layout in space is linear and ordered left to right, top to bottom. In terms of the visual nature of the individual elements and their spatial arrangement, none of these levels of descriptive representation directly maps onto the visuospatial structure of the subject matter it represents. Descriptive representations can therefore be thought of as *arbitrary* in the sense that there is essentially no discernible one-to-one correspondence between the representation and its referents [2]. However, mental models supposedly represent their referents in a far more analog fashion. The external subject matter that is modelled internally via these knowledge structures is represented by tokens (mental entities that stand for the external referent entities) that are arranged in a relational organization paralleling the key referent relationships. In other words, a mental model is partly isomorphic in that it has a high degree of correspondence at a fundamental level with the referent subject matter it represents. It follows that a person engaged in reading a piece of text about content that is markedly visuospatial in nature must carry out extensive conversion of that representation in order to process it into a form that is well suited for mental model construction. This transformational side-task can be a very resource intensive process in its own right and hence tie up capacity that could otherwise be devoted to the main task of mental model construction. Table 2 summarizes some types of conversion activities that a reader of text may need to carry out during this transformational processing.

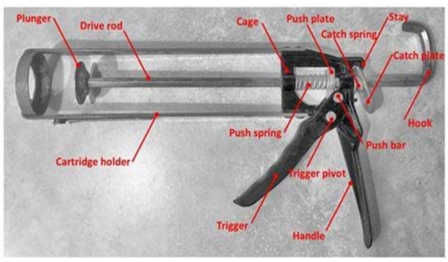
Identify all task-relevant entities referred to (directly or indirectly) by the text	**Mentally organise tokens into an arrangement analogous to the structure of the referent situation**
Determine the material properties of all relevant entities	**Assign task-relevant dynamic information to mental tokens**
Determine the geometric properties of all relevant entities	**Introduce appropriate constraints on tokens' dynamics**
Convert non-analog (linearized) account of visual and spatial information into analog visuospatial account	**Form sequences of functionally-related tokens to represent causal chains**
Abstract information about entities to generate appropriate mental tokens	**Incorporate higher-order relations governing the coordination of different causal chains**

Table 2. Text-to-visuospatial conversion activities

In contrast, the visuospatial properties of a well-designed *depictive* representation of the same content already closely match those of its referent subject matter. It typically has a high degree of one-to-one correspondence to its referent in terms of both the entities it portrays and how those entities are related to one another. Unlike text, there are few arbitrary conventions (such as linearity and sequencing rules) that intrude to distort the layout of the represented information. Instead, in all but the most extreme cases of abstraction and manipulation in diagrammatic depictions, mapping between representation and referent is relatively straight forward. This means that the task of going from a depictive representation to a mental model is likely to involve far less side-task transformational processing than would be the case for a corresponding descriptive representation. In the next sections, we expand on how differences in descriptive and depictive representations have a fundamental effect on the way they are processed in a multimedia context.

2.3 Reconciling representations: Processing Implications

Consider a conventional multimedia resource that presents complex, unfamiliar subject matter via a text and an accompanying overview picture. The text shown in Figure 2 addresses the 'push' subsystem of the caulking gun mechanism using such an approach.



Conventional multimedia text component to accompany above picture

For each squeeze of the caulking gun's trigger, the following cycle takes place: When the trigger is first depressed, it initially rotates on its pivot and moves part way toward the handle. This causes the push bar to push on the bottom of the push plate which swings to the left. As a result, the push plate changes from being perpendicular to the drive rod to being angled to that rod. As depression of the trigger continues, the push plate then begins to compress the push spring as it pushes the drive rod to the left. This moves the plunger along inside the cartridge holder. Further depression of the trigger continues movement of the drive rod and plunger through the cartridge holder until the push spring is fully compressed against the side of the cage and the trigger is stopped by contact with the handle. When the trigger is released, the trigger, push plate and push spring return to their original states.

Fig. 2. Multimedia presentation of caulking gun subsystem

A person encountering this combination must make many back-and-forth comparison transitions between the text component and the picture component while trying to build a coherent unified mental representation of the referent system. This involves repeated shifts of attention and attentional adjustments in order to process corresponding or complementary target aspects of these two very different types of representation. In addition to navigating these transitions, the individual must also perform various mental conversions in order that the information carried by these two very different media can be reconciled and then combined on a common representational basis (see Table 2). Both the continual to-and-fro activity and the ever-present requirement for such conversions are resource-intensive processing activities. For subject matter of any complexity, they result in this being a very demanding form of processing, both perceptually and cognitively.

Now consider how the situation (and concomitant processing demands) would likely change if the text was replaced by ancillary diagrams depicting the same content as would normally be presented by that text. We hypothesise that although some back-and-forth comparison transitions would still be needed, fewer of them should likely be required and they would impose a much smaller processing overhead. This is because of the far greater similarity between the diagrams and the picture than between text and picture. As a result, search within the diagrammatic depictions would be substantially lower than within the body of text and would be less demanding. Then, once corresponding representations of the same information items were located, much less processing would be required to reconcile them.

Keeping in mind the limits on human information processing, the second scenario should be considerably less demanding and therefore likely to leave more processing resources free for mental model building than would the text-based scenario. Further, because any inter-representational conversions that are required would be far more modest, there should be less danger of errors creeping in with the diagram-based scenario.

3 Constructing a Mental Model: Diagrams Instead of Text?

In this section, we consider in more detail the task of constructing a mental model from an overview picture accompanied by ancillary diagrams (rather than text). To assist our consideration of this task, we devised Figure 3 as a notional way to show some key features of a mental model in concrete form. Note that we do not claim Figure 3 to be anything more than a hypothetical expression of these features. It is intended solely to facilitate comparison between a mental model and the type of ancillary diagrams discussed above. This is an important point to make because there is currently no definitive account of how mental models are actually manifested in the mind. Johnson-Laird [6,7,10] considered them to be analog representations that preserved structural aspects of the referent and although grounded in perception, were abstract and not modal-specific. Their abstract nature means that rather than representing *particular* situations (the case with mental images), mental models represent *sets* of situations.

Despite Figure 3 being a much simplified representation of the caulking gun that lacks veridical detail, it nevertheless bears a close structural resemblance to the gun's mechanism. Although it preserves the fundamental relationships between the entities that exist in a real caulking gun, abstract tokens are used in place of the referent's entities. Further, the representation singles out the functional role of key aspects (e.g., the pivots) and captures certain properties that are relevant to how the device operates (e.g., the tokens representing entities that are fixed versus those that are moveable). For the sake of clarity, Figure 3 does not attempt to be comprehensive - other aspects could be added that would probably bring it closer to a 'real' mental model (such as information about the direction and extent entities can move, sequencing information, etc.). However, the purpose of this realization is merely to help demonstrate that using ancillary diagrams in multimedia resources may be a more effective way to support mental model construction than the conventional use of text.

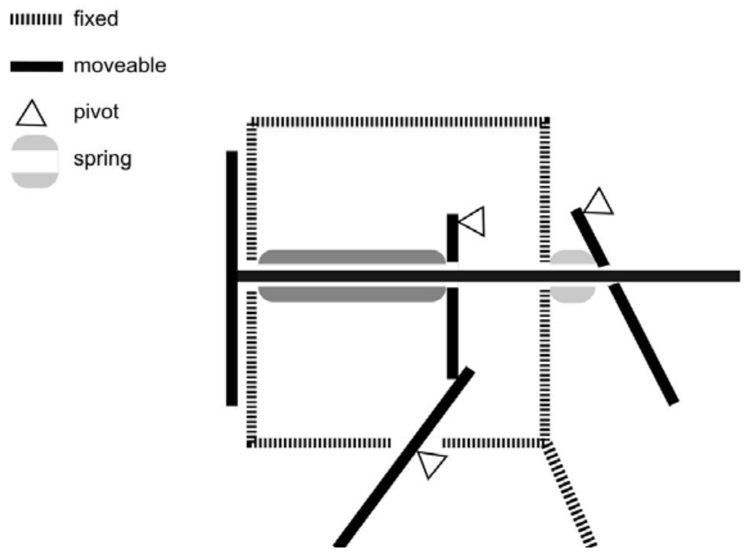


Fig. 3. Possible information in a mental model of the caulking gun mechanism (hypothetical)

Let us assume that Figure 3 is a not unreasonable concrete expression of what an individual is trying to construct when processing a multimedia resource on the caulking gun mechanism. We can then use Figure 3 as a basis for hypothesizing about the processing routes and activities that might be involved if an individual was to be given ancillary diagrams (instead of text) as a basis for mental model construction. The type of situation envisaged here is that the block of text provided in Figure 2 would be replaced by several diagrams intended to serve the same content presentation purpose. Figure 4 shows a possible implementation of such a combination.

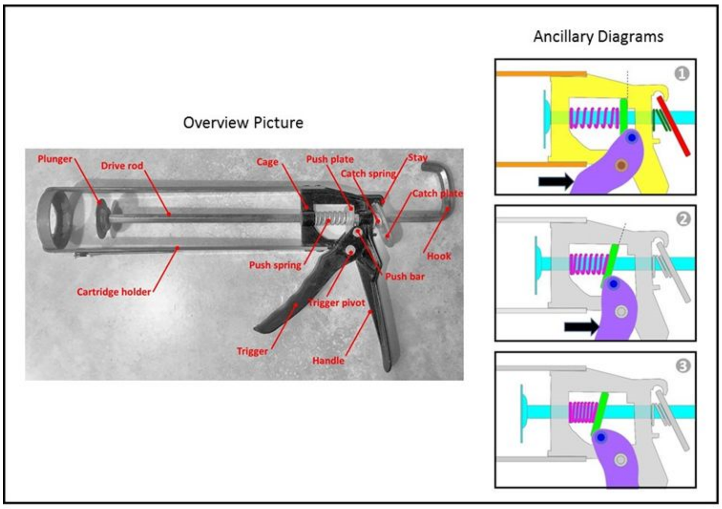


Fig. 4. Ancillary diagrams used in place to text to accompany overview picture

Note that the ancillary diagrams incorporate various features that are designed to boost their explanatory power (e.g., colour coding, transparency, additional symbols, alignment to aid inter-diagram comparisons, etc.). For comparison purposes, Figure 5 places the first ancillary diagram beside the information set constituting the hypothetical mental model.

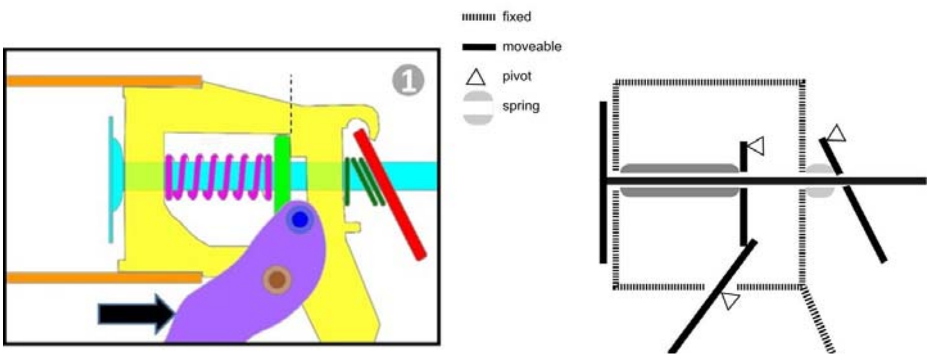


Fig. 5. Ancillary diagram/mental model comparison

Although they are various superficial differences between the two representations, at a deeper structural level they have a great deal in common. This makes it relatively easy to map between their corresponding aspects. It is also quite possible to do similar mapping with the other two ancillary diagrams shown in Figure 4 if the mental model is progressively ‘run’ forwards in time. For example, the relation between the solid line token representing the trigger and its adjacent pivot token in the mental model allows domain general information about how a lever rotates when a force is applied to one of its arms to predict how the push bar will move when the trigger is pressed.

However, for the situation being addressed here (i.e., constructing an internal representation from external ones), the circumstances are reversed. In that case, the person studying the overview-plus-ancillary diagrams composite would be using those depictions to abstract a representationally efficient generalization about this type of behavior in order to incorporate it into a developing mental model. We can envisage this would involve comparing the three successive ancillary diagrams and inferring that the trigger can move smoothly towards the handle by rotating around its pivot. The mental model would be constructed to represent this aspect of the caulking gun's functionality in a parsimonious manner as a continuous process (rather than as a series of discrete stages as shown in the ancillary diagrams). During this mental model construction activity, instance-specific information shown in the ancillary diagrams that would tend to limit the potential generalizability (and hence power) of the developing representation would presumably be omitted. For example, aspects such as the cartridge holder, the particular shaping of the trigger and handle, the small hook that constrains the top end of the catch plate, etc., would either be dispensed with completely or tokenized to maximize their generalizability. This would mean that the mental model could be applied not only to the particular instance of a caulking gun shown in Figure 1, but also to a host of other superficially different but functionally similar design variants of this device that are available in the market place.

4 Repurposing multimedia's text components

If empirical studies were to show that replacing text with ancillary diagrams did indeed improve the understandings developed from multimedia explanations, does this mean that text would then be absent from such resources? Our view is that this should not be the case. Rather, we suggest that text's traditional role of presenting the content could be replaced by a new and very different role. This alternative role would be to guide users of a multimedia resource in how to optimize their interactions with the ancillary diagrams (in coordination with the overview picture). Research has shown that if pictorial materials are presented without sufficient guidance as to how they should be interrogated, interpreted, and inter-related, understanding of their contents may be compromised [11,12]. This potential deficiency can be related to the fact that, unlike text, there is a lack of standardized reading conventions and approaches that can be applied across all instances of pictorial representation.

Well written text, irrespective of the topic, leads the reader systematically through the presentation of its subject matter by taking advantage of its standardized linear structure and syntactical rules. However, there are no corresponding constraints on how pictures are to be 'read' because the presented information is structured according to the structure of the depicted subject matter (and not according to a universally applicable set of conventions and rules). As a consequence, individuals who lack background knowledge about the subject matter portrayed in a depictive representation may prioritize which aspects they attend to and how they sequence their interrogation of those aspects according to the depiction's superficial perceptual characteristics [c.f. 13]. This may result in key information going un-noticed and errors of interpretation.

Because depictive representations such as the ancillary diagrams being addressed here lack the inbuilt features for supporting appropriate navigation etc. that are present for text, it would make sense to accompany them with some form of add-on guidance. In contrast to text's limitations for representing visuospatial information, it can be a most effective way to convey sequenced procedural instructions. Our suggestion therefore is that rather than removing text from multimedia resources altogether, it instead be repurposed as a way of guiding the viewer through appropriate and fruitful processing of the ancillary diagrams that we suggest might take over text's traditional role. An illustration of how such an approach might be implemented is given in Figure 6

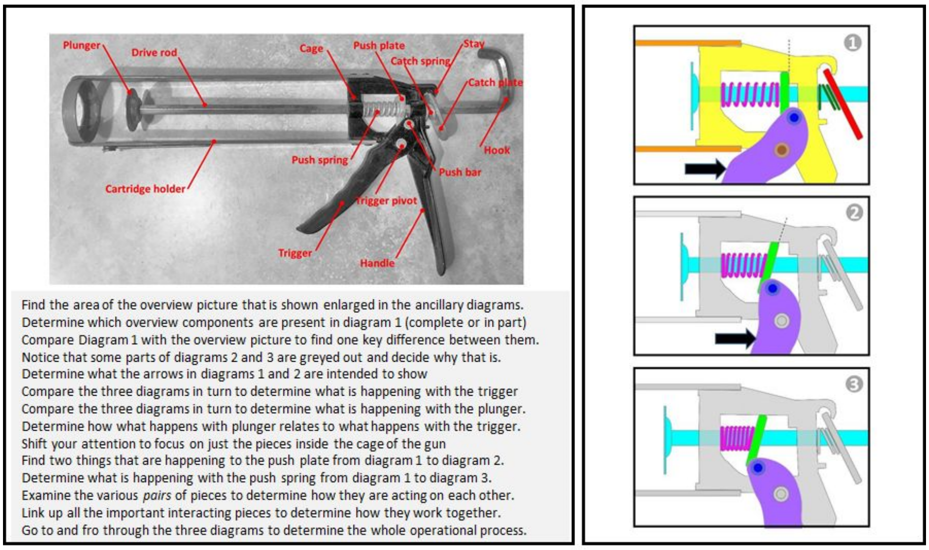


Fig. 6. Guiding text could support interrogation of depictive representations

Another justification for changing the role of text to guiding interrogation of ancillary diagrams (instead of presenting content) is that this could largely avoid the issue of requiring possession of domain-*specific* background knowledge (almost unavoidable when text is used to present content). Most text-based explanations of content that is in some ways complex or unfamiliar implicitly assumes a certain existing level of relevant background knowledge (often wrongly, which hinders the reader's interpretation). It is really very challenging for a text author to provide for a range of readers who have widely differing prior knowledge of relevance to the particular content involved.

Converting from a text explanation of some content to an analog representation typically requires the person to elaborate the entities mentioned. So, if the text mentions a spring, the reader will need to 'flesh out' that item in order to convert it... such as what shape springs usually are, what they are made of, how they behave when subjected to force, their ability to return to their original shape/size once the force is removed, etc. A person who lacks such 'spring-specific' background knowledge could be at a real disadvantage with a text-based explanation of our caulking gun mechanism.

If text was instead used for guidance, not content presentation, it would rely only on domain *general* background knowledge, such as what is meant when someone is asked to focus attention on a specific areas of a diagram, notice a particular aspect of that diagram, or compare two of its aspects, etc. These guiding instructions are extremely generic and are universally applicable, irrespective of the content involved. It would be almost unheard of that someone would not understand what they were required to do having received such an instruction. Even if guiding text introduced some unfamiliar type of interrogation activity, it would be easy to explain what was intended without relying on the person having specialist background knowledge.

5 Ancillary diagrams: animated alternatives

Considering the practicalities of using a set of ancillary diagrams instead of text to accompany an overview picture of the referent subject matter raises some potential limitations of this approach. One of these limitations arises from the static nature of such diagrams (which means that relevant dynamics must be portrayed indirectly rather than directly). Representing dynamics via static diagrams requires extra information over and above that necessary to depict only visuospatial aspects of the referent subject matter [12]. For example, to convey information about changes in components over time, multiple ancillary diagrams of the caulking gun mechanism were used in Figure 4. This approach addresses such changes for the trigger, push plate, push spring, drive rod, etc. Further, within-diagram additions are also necessary, such as the inclusion of arrows (to show the force applied to the trigger that causes it to move) and the dotted line to indicate the change in orientation of the push plate). These extras not only increase the number of depictions that a viewer must deal with but also result in those displays becoming more cluttered. Both types of addition can therefore raise the processing demands imposed on viewers. Another potential downside of these static ancillary diagrams is that the viewer is required to correctly interpret their extra dynamics-related information. This interpretation relies on the viewer possessing and successfully applying appropriate background knowledge about the conventions used to indicate dynamics via a static depiction. For viewers who are young or who lack such knowledge, this requirement can result in difficulties and interpretation errors.

Using animations instead of static diagrams may offer a way to avoid these potential problems. It could reduce or eliminate the need not only for multiple diagrams, but also for within-diagram additions. Further, it provides an opportunity to make explicit the relationship between the information in an overview depiction and in its accompanying ancillary diagrams. For example, animation could be used to single out a target sub system from the overview depiction and convert it into an ancillary diagram. It could also be used in a related way to show the origin of ancillary diagrams that provide views of the subject matter that are different from those given in the overview depiction.

However, research indicates that animations are by no means a universal panacea. Ironically, the very strength they have in terms of being able to represent dynamics directly can be problematic when it comes to viewers extracting task relevant information from animated displays [14]. Avoiding such problems seems to rely on fundamentally changing how animations are designed so that the way they present information is more closely attuned with human information processing capacities. The Composition Approach to animation design has been developed to address the mismatch between these two aspects that too often compromises the effectiveness of conventionally designed ('Comprehensive') animations. This novel approach is founded on the Animation Processing Model (APM) that characterizes the construction of a mental model from an animation in terms of five interdependent processing phases [15].

If animated rather than static depictions were to be used as ancillary diagrams in the approach canvassed above, it is important that their design be optimized in terms of providing support for mental model construction. One important consideration in designing more effective animations is the fine-grained, content-specific characterization of likely perceptual and cognitive challenges that could arise from presenting information about the referent subject matter in a temporally veridical manner. Complex, unfamiliar dynamic subject matter tends to have features that makes it difficult for viewers to process successfully if an animated representation faithfully reproduces its actual dynamics. This typically occurs if the dynamics involve substantial simultaneity, as is the case with the caulking gun example. Here, the 'push' causal chain and the 'catch' causal chain have an intimate functional relationship (the gun will not work as it should unless both operate properly and in concert). However, if an animation was actually to depict these two central aspects of the gun simultaneously (i.e., as they would occur in real life), a likely result would be inadequate viewer processing of the presented information. As a consequence, the quality of a mental model built from exposure to this animation would probably be severely compromised.

Empirical research indicates the benefits of a Composition Approach to animation design in which counterproductive simultaneity is removed by sequential rather than parallel presentation of such information [16,17]. A necessary pre-cursor to this design approach is a thorough analysis of the referent subject matter that can reveal which of its aspects have the potential to impose excessive information processing demands on the viewer if presented in a veridical manner. The results of such an Event Unit Analysis [18] can them be taken into account in the design of an animation to tailor its presentation characteristics to those of the particular subject matter being addressed. Event Unit Analysis is therefore a content-specific technique whose results will differ according to the nature of the subject matter involved.

Discussion and conclusion

A dominant concern of orthodox approaches to multimedia design is to find ways of staging presentation of their constituent descriptive and depictive representations so that their combination works as effectively as possible. Typically however, these approaches do not fundamentally question the roles that are assigned to these two broad types of representation within such combinations. In this paper, we considered the potential effectiveness of multimedia resources from the perspective of the extent to which their design was likely to provide support for an individual's construction of high quality mental models. From an analysis of their likely processing demands, we concluded that descriptive representations (e.g., text) should be far less suited to conveying content information than depictive representations. Using the example of a caulking gun mechanism, we explored an alternative approach to multimedia design that replaced explanatory text with ancillary diagrams and repurposed text as a guiding resource for supporting more effective processing of the depictive content presentation. Potential limitations of this approach were flagged and a suggestion made that they could be circumvented by using animated instead of static ancillary diagrams [19]. Cautions about the possibility of negative effects from conventionally designed animations were raised and the use of Composition Approaches and Event Unit Analysis introduced as a way to ameliorate these effects. The theoretical proposals put forward in this paper are intended to stimulate empirical research into more principled ways to design explanatory multimedia. Suitable experiments could range from straightforward comparisons of the relative effectiveness of text versus ancillary diagram accompaniments in multimedia, to the potential of more tokenized animated ancillary diagrams for fostering higher quality mental models [20]. However, we recommend that this research gives more consideration to the benefits of 'working backwards' from posited attributes of high quality mental models rather than merely 'working forwards' from the characteristics of the external representations that supply the raw material from which those internal representations are built.

References

1. Wright, P.: Presenting technical information: A survey of research findings. Instructional Science, 6, 93-134 (1977).
2. Schnotz, W.: Integrated Model of Text and Picture Comprehension. In R.E. Mayer & L. Fiorella (Eds). The Cambridge Handbook of Multimedia Leaning, pp. 82-99, Cambridge: Cambridge University Press (2020).
3. Butcher, K.R.: The Multimedia Principle. In R.E Mayer (Ed), The Cambridge Handbook of Multimedia Learning (2nd Edition), pp. 174-205, Cambridge: Cambridge University Press (2014)
4. Low, R., Sweller, J. The Modality Principle in Multimedia Learning. In R.E Mayer (Ed), The Cambridge Handbook of Multimedia Learning (2nd Edition), pp. 227-262, Cambridge: Cambridge University Press (2014)
5. Kintsch, W.: Comprehension: A paradigm for Cognition. Cambridge: Cambridge University Press (2003).

6. Johnson-Laird, P.N.: Mental Models. Cambridge, MA: Harvard University Press (1983).
7. Johnson-Laird, P.N.: How We Reason. Oxford: Oxford University Press (2006).
8. Bétrancourt, M.: The animation and interactivity principles in multimedia learning. In R.E. Mayer (Ed), The Cambridge Handbook of Multimedia Learning (1st Edition) pp. 487-512. Cambridge: Cambridge University Press (2005). Doi: 10.1017/CBO9780511816819.019.
9. Lowe, R.K., Boucheix, J.M. & Menant, M.: Perceptual processing and the comprehension of relational information in dynamic diagrams. In P. Chapaman, G. Stapleton, A. Moktefi, S. Perez-Kriz, F. Bellucci (Eds), Diagrammatic Representation and Inference, pp.470-483. Proceedings of the Diagrams Conference, LNAI 10871, Springer Nature (2018).
10. Johnson-Laird, P.N.: Mental models and human reasoning. PNAS, 26, 107 (43) 18243-18250 (2010). https://doi.org/10.1073/pnas.1012933107
11. Hegarty, M., Just, M.A.: Constructing mental models of machines from text and diagrams. Journal of Memory and Language, 6, 32, 717-742 (1993).
12. Mayer, R.E., Hegarty, M., Mayer, S., Campbell, J.: When static media promote active learning: annotated illustrations versus narrated animations in multimedia instruction. Journal of Experimental Psychology: Applied, 11, 4, 256- 265 (2005).
13. Lowe, R.K., Boucheix, J.M.: Learning from Animated Diagrams: How are Mental Models Built? In, G. Stapleton, J. Howse, and J. Lee (Eds.): Diagrams 2008, LNAI 5223, pp. 266–281, Springer-Verlag Berlin Heidelberg (2008).
14. Lowe, R.K.: Animation and learning: Selective processing of information in dynamic graphics. Learning and Instruction. 13, 157-176 (2003).
15. Lowe, R.K., Boucheix, J.M.: Dynamic Diagrams: A composition Alternative. In P. Cox, B. Plimmer, & P. Rogers (eds.), Diagrammatic Representation and Inference, LNAI, pp. 233-240, Berlin: Springer (2012).
16. Lowe, R.K., Boucheix, J.M.: Principled animation design improves comprehension of complex dynamics. Learning an Instruction, 45, 72-84 (2016). https://doi.org/10.1016/j.learninstruc.2016.06.005.
17. Lowe, R.K., Schnotz, W. & Boucheix, J.M.: The Animation Composition Principle in Multimedia Learning. In Mayer, R.E., & Fiorella, L. (Eds.), The Cambridge Handbook of Multimedia Learning (3rd Edition.) pp. 313-323. New-York: Cambridge University Press. (2021).
18. Lowe, R.K., Boucheix, J.M.: Event unit Analysis: A methodology for anticipating processing demands of complex animated diagrams. In, A.-V. Pietarinen, P. Chapman, L. Bosveld-de Smet, V. Giardino, J. Corter, S. Linker (Eds.): Diagrams 2020, LNAI 12169, pp. 307–322, Springer Nature (2020). https://doi.org/10.1007/978-3-030-54249-8_24.
19. Ploetzner, R., Berney, S., Bétrancourt, M. A review of learning demands in instructional animations: The educational effectiveness of animations unfolds if the features of change need to be learned. Journal of Computer-Assisted Learning, 36, 838-860 (2020). https://doi.org/10.1111/jcal.12476
20. Boucheix, J.M. Lowe, R.K.: 'Tokenized' dynamic diagrams: An approach for improving mental model construction? In, A.-V. Pietarinen, P. Chapman, L. Bosveld-de Smet, V. Giardino, J. Corter, S. Linker (Eds.): Diagrams 2020, LNAI 12169, pp. 481–484, Springer Nature (2020). https://doi.org/10.1007/978-3-030-54249-8_24.