Title

The Pragmatics of Explanation

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN

1069-7977

Authors

Chin-Parker, Seth
Bradner, Alexandra

Publication Date

2008

Peer reviewed

The Pragmatics of Explanation

Seth Chin-Parker (chinparkers@denison.edu)
Department of Psychology, Denison University
Granville, OH 43023 USA

Alexandra Bradner (bradnera@denison.edu)
Department of Philosophy, Denison University
Granville, OH 43023 USA

Abstract

The pragmatic theory of explanation (Van Fraassen, 1988) proposes that background knowledge constrains the explanatory process. Although this is a reasonable hypothesis and research has shown the importance of background knowledge when evaluating explanations, there has been no empirical study of how the background constrains the generation of explanations. In our study, participants viewed one of two sets of preliminary movie clips of some novel items engaged in a series of actions and then all were asked to explain the same final clip. Between conditions, we varied whether the events in the preliminary clips completed a system. In the systematic condition, a greater proportion of functional explanations were generated for the final clip compared to the non-systematic condition. Interestingly, despite the difference in the types of explanations generated, the participants showed high agreement in the evaluation of explanations provided by the experimenters.

Keywords: explanation; cognition; function; pragmatics.

Explanation holds a special place in the cognitive sciences (Lombrozo, 2006). Philosophers, psychologists, members of the artificial intelligence community, and computer scientists have all studied various facets of explanation. The allure to cognitive scientists is obvious: An explanation embodies an individual's ability to express the understanding that she has acquired for events that occur in the world. If she is able to explain why peaches are fuzzy or why a person shivers, it indicates that she has found meaning in the relationship between those simple facts (a peach is covered in fuzz) or events (the person is shivering) and her more general knowledge of the world. The explanation goes beyond describing the event or recognizing associations, e.g. it is cold and that person is shivering, to providing connections between the explanandum (what is to be explained) and the explanans (what does the explaining).

Philosophical inquiry into the nature of explanation has a long and rich history. At the risk of over-simplifying this body of work, we propose that there have been three main approaches to what constitutes an explanation. Much work has illuminated our understanding of causal explanation (e.g. Salmon, 1998), functional or teleological explanation (e.g. Cummins, 1975; Wright, 1973), and the unification approach to explanation (e.g. Kitcher, 1981). In the current study, we are interested in the cognitive underpinnings of explanation as opposed to the explication of an explanation itself, so instead of focusing our attention on one of these approaches exclusively, we instead explore a fourth approach, the pragmatic approach, as proposed by Van Fraassen (1988). This approach subsumes the other three, for explanation is seen as a selection process (Bradner, 2005) – at any time there are multiple candidate explanations that could be referenced, but we choose the most appropriate for a given situation by working through two stages. First, the topic of the explanandum is determined by identifying the contrast class for the explanation, defining the explanandum. For instance, the answer to "Why is *she* shivering – as opposed to why is *he* shivering?" is different than the answer to "Why is she *shivering* – as opposed to why is she *dancing*?" After the contrast class has been identified, relevance relations motivate the content of the explanation. The relevance relations for "Why is she shivering?" would possibly include knowledge of human physiology or knowledge of the temperature outside. Van Fraassen proposes that background knowledge does a good deal of the work in constraining the possible explanations at each of these stages. Earlier work in artificial intelligence (e.g. Leake, 1991) has explored the implications of this view, but there has not been empirical study of how background knowledge affects the selection of the contrast class or the relevance relations.

Recently, there has been active study of what parameters influence when causal explanations (e.g. Glymour, 1998) and functional explanations (e.g. Chaigneau, Barsalou, & Sloman, 2004; Lombrozo & Carey, 2006) are warranted. In the current study, we extend this work by asking participants to provide an explanation for an explanandum when either a causal or functional explanation is licensed. By manipulating the prior experiences of the participants, we are then able to assess whether the pragmatic theory is correct in predicting that the participants' explanations are determined by available background.

Although there are different proposals as to how people come to appreciate the presence of causal relations (see Shanks, Holyoak, & Medin, 1996, for one overview), there is consensus that cause is often invoked during explanation (Keil, 2006). We expected that participants would readily generate causal explanations for events that were physically and temporally contiguous and that followed basic beliefs associated with naïve physics – e.g. the movement of object A as object B moves into a space previously held by object A can be explained by object B causing object A to move.

The research examining functional explanations suggests that something more than an appreciation of causation is necessary for an individual to understand the function of an item. Lombrozo and Carey (2006) adopted, and found evidence supporting, a view of functional explanation that was firmly rooted in the work of Wright (1973). According to this view, functional explanations are licensed when the history of the explanandum indicates that it was intended to fulfill that function. For example, the function of pumping blood provides an explanation for the presence of the heart. An alternative perspective is offered by Cummins (1975). He proposes that the function of an item can only be understood in reference to the role the item plays in a larger system. According to this view, function is not what anachronistically explains the presence of something but what results from a series of items fulfilling their capacities within a larger causal structure. Chaigneau, Barsalou, and Sloman (2004) tested the relative contributions of the history and causal structure to an appreciation of function in a series of experiments. Within the framework of their HIPE theory, they posited that an item would be identified with a particular function either by having been made to fulfill that function, a historical account in line with Wright (1973), or by having an appropriate causal structure for the function, as suggested by Cummins (1975). Their study showed that knowing the history was sufficient only if the immediate causal structure was unknown or there were no contradictions to the intended use in what was known. If the causal structure allowed for particular function, participants rated the function as present even if the item was not licensed that function by its history.

We adopt a notion of function similar to Cummins (1975), because it allows us to make a distinction between a causal explanation, which explains the relationship between items without reference to a larger system or purpose, and a functional explanation, which might invoke those same causal relationships but does so while pointing to the purpose within a larger system. For instance, a causal explanation provides some information that goes beyond a simple description of two related events, providing some backdrop so that "X *then* Y" can be appreciated as "X *caused* Y." A functional explanation goes a further step in that the relationship between X and Y can be appreciated in terms of some larger system, "X *caused* Y *in order that* Z." Based on these ideas, we propose that prior exposure to events that occur within a complete system will prompt participants to generate a functional explanation. Exposure to the same events without the structure afforded by a complete system will point participants towards a more localized, causal explanation for an event. We test this hypothesis in the current experiment by having participants explain an animated clip that shows a series of events and they are asked to explain what they saw. Prior to seeing this clip, the participants see some preliminary movie clips that are similar to the final clip, and we manipulate whether the events within those clips complete a system in the sense that the initial event causes events that lead to the final event and the final event is in turn related to a recurrence of that initial event. We predict that participants that see the systematic preliminary clips will come to appreciate the larger causal structure of the events and they will be more likely to generate functional explanations for the final clip. The participants that see the non-systematic clips, ones where there is a break in the larger systematic relationship of the events, will instead generate causal explanations since their prior experiences would not have emphasized the larger system. This is the first empirical study we know of that aims to analyze explanations generated by the participants. Since prior work has focused on the evaluation of explanations, we also include an explanation selection task. For that measure, we predict that participants in the systematic condition will be more likely to select a functional explanation while participants in the non-systematic condition will tend to select a causal explanation.

Experiment

Methods

Participants Thirty-three undergraduates from a Midwest college participated in this experiment. They completed the experiment following participation in an unrelated problem solving study and were compensated for their participation. One participant's data from the selection task were lost due to experimenter error. Three participants' generation data were not available for analysis because the microphone was not turned on for two sessions, and one participant did not follow the directions to verbalize his explanation before moving onto the subsequent selection task.

Design Participants were randomly assigned to either the systematic condition or the non-systematic condition, and this assignment determined which set of preliminary movie clips the participant viewed. The order of the preliminary movie clips was randomized for each participant. All participants completed an explanation generation task and an explanation selection task after viewing the preliminary movie clips. All participants completed the generation task before the selection task, and the order of the explanations in the selection task was balanced across participants.

Materials The primary materials for the study were short animated movies. In each of the preliminary clips, a colorful, oddly-shaped object pushed onto a round object that compressed, causing a lever to move. The lever was connected to a container of some sort by means of other devices, and when it moved, the container released some of what it was holding. In the *systematic condition* clips, the released material fell into an opening in a segment of the oddly-shaped object (see Figure 1a and 1b). This final event completed the system in that the oddly-shaped object began the events by pushing and final result of that action was a return to the oddly-shaped object. In the *non-systematic condition* clips, the exact same events occurred except that the material released piled up next to the oddly-shaped object (see Figure 2a and 2b). In this case, the larger system

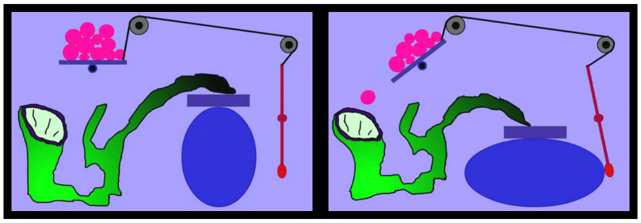

Figure 1a Figure 1b

Figure 1: Initial scene (1a) and a midpoint scene (1b) of one of the preliminary clips from the systematic condition.

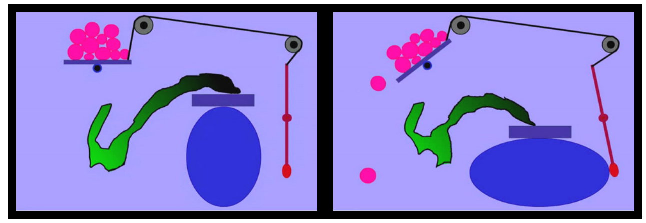

Figure 2a Figure 2b

Figure 2: Initial scene (2a) and a midpoint scene (2b) of one of the preliminary clips from the non-systematic condition.

is compromised since the final result does not return the action back to the oddly-shaped object. Otherwise, the preliminary movie clips seen in the two conditions were identical. In each of the clips, the series of events was repeated three times. The final clip that was shown to all participants was much like the initial clips – an oddly-shaped object pushed onto a round object and, through a series of events, some material was released. However, the final movie clip was intended to be ambiguous as to the fate of the released material. It spilled near to the oddly-shaped object, as in the non-systematic clips, but there was a segment of the oddly-shaped object near to where it collected (see Figure 3).

For the explanation selection task, we prepared four possible explanations for the final clip; one causal explanation, one functional explanation, one description of the objects involved in the final clip, and one inaccurate description of the objects and events in the clip. The explanations used in the task can be found in the Appendix.

Procedure All participants were told that they would be seeing a series of movie clips and then they would be asked to provide an explanation for a final clip. Participants saw the preliminary movie clips; after the third clip, they saw a screen for ten seconds that said, "Take a moment to think about the movie clips you have just seen." That screen was replaced by instructions that notified the participants that they would see one last movie clip two times, and then they would be prompted for an explanation. After the final movie clip played twice, the participants were instructed: "Please provide a possible explanation for what you just saw in the previous clip." This *explanation generation* task was recorded for later analysis. Once the participants finished their explanations, they moved onto the *explanation selection* task. In this task, the participants were provided with four possible explanations for the final clip. The participants were asked to click on the "best" explanation. The explanation chosen disappeared, and the participants were asked to again choose the best explanation from the remaining selections. This was repeated a third time, providing a measure of what the participants considered the best to worst explanations (the worst being the unselected explanation). The participants also completed a rating task where they rated an extensive list of possible explanations, but the measure was not a critical one for the current study, so it is excluded in the interest of brevity[1].

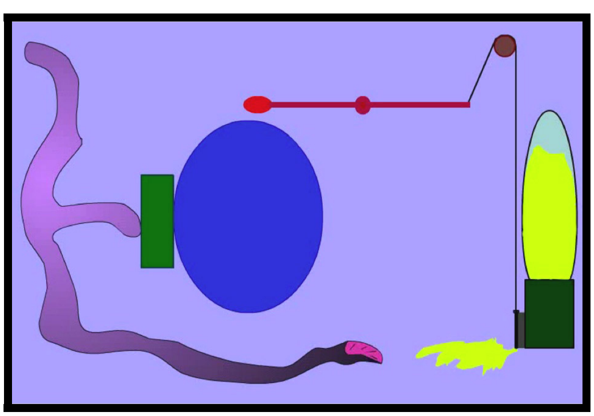

Figure 3: Final scene from the clip seen by all participants, regardless of condition. The participants were asked to explain what they saw in the clip.

Results

Our primary interest in this study was whether the preliminary clips influenced the type of explanation generated by the participants. We classified the explanations as causal, functional, or neither. In order for an explanation to be considered causal, it had to make use of the causal structure described above, "X caused Y". We accepted several similar phrases in place of "caused", such as "triggered," "made happen," or "resulted in" to satisfy the causal distinction. Functional explanations had to go beyond identifying causal relationships and provide some larger purpose to the actions, "X caused Y in order that Z". As was noted prior, the appreciation of the purpose can diminish the necessity of the immediate causal precursors in a functional explanation, so we focused on "in order that Z" in identifying functional explanations. Thus, for an explanation to be identified as a functional explanation, it had to reference a larger purpose for the system.

In order to analyze the *explanation generation* measure, the two authors independently coded the explanations that

[1] A subset of the ratings was related, but the pattern they revealed was very similar to the explanation selection responses, so they provided no new information with regards to the current study.

the participants generated, agreeing on 24 of the 30 explanations – 80% inter-rater agreement. The other six explanations were revisited, and the authors easily came to a consensus on the appropriate coding for those explanations. Examples of the explanations are included in the Appendix.

Of the 15 explanations from the non-systematic condition, 13 were coded as being causal and 2 were coded as being functional. Of the 15 explanations from the systematic condition, 3 were coded as being causal explanations and 12 as functional explanations. The distribution of the explanations across the conditions shows a strong effect of the preliminary clips, $\chi^2_{(1, 30)} = 13.39$, $p < .01$, $\phi_C = .67$. As predicted, the participants in the systematic condition showed a greater proportion of functional explanations than the participants in the non-systematic condition. When reviewing the explanations, we noticed two other patterns that differentiated the conditions. In the systematic condition, seven of the participants described the oddly-shaped object as a living thing while in the non-systematic condition only two participants did the same, $\chi^2_{(1, 30)} = 3.97$, $p < .05$, $\phi_C = .36$. Also, in the systematic condition, nine of the participants incorporated the idea that the final clip illustrated a failure of a system, while none of the participants in the non-systematic condition did the same, $\chi^2_{(1, 30)} = 12.86$, $p < .01$, $\phi_C = .65$. These patterns were not predicted a priori, but they are informative in terms of understanding how background knowledge was used in the explanatory process.

The pattern of results found in the *explanation selection* measure showed a clear preference for the causal explanation followed by the functional explanation, and there was no difference between the groups. Thirteen of the sixteen participants in the systematic condition chose the causal explanation first, and the other three in that condition chose the functional explanation initially. In the non-systematic condition, fourteen of the sixteen selected the causal explanation, and the other two selected the functional explanation initially. Nearly every participant that had chosen the causal explanation first chose the functional explanation second. The five participants who had chosen the functional explanation first all chose the causal explanation second. Also, in both groups, an accurate description of the objects in the final clip was selected before an inaccurate description of the events in that clip showing that the participants were able to remember what had happened during the final clip while they were doing the task. The lack of a difference between the conditions' selection patterns was counter to our predictions.

The presence of an effect of the preliminary clips in the generation task but not the selection task suggests a dissociation between the measures. In order to test this possibility, we coded each participant's response for each measure as either consistent or inconsistent with the predictions that participants in the systematic condition would generate and select functional explanations while participants in the non-systematic condition would generate and select causal explanations. For this analysis, we were able to use the data from the 29 participants from whom we had both the generation and selection data. In the generation task, 24 of the 29, or 83% of the participants, generated explanations that were consistent with our predictions. In the selection task, 15 of the 29, or 52% of the participants, selected explanations in a manner consistent with our predictions. Using a McNemar's test, we found a significant relationship ($p = .01$) between the type of task and the proportion of the participants responding in a manner that was consistent with our predictions.

Discussion

The results of the experiment support the basic premise of the pragmatic approach to explanation. For any given explanandum, there are multiple candidate explanations, and the background of the explainer plays a critical role in determining what explanation is generated. However, we also found an unexpected difference between the results of the explanation generation and selection measures.

First, we examine the differentiation between the two conditions on the explanation generation task. Although the results are in line with the predictions we made a priori, the content of the explanations indicate that the explanatory patterns we observed are the result of two distinct effects of the background knowledge. First, as we noted in the Results, a majority of the participants in the systematic condition made reference to the failure of a system in the final clip. Their sensitivity to the presence of a larger system in the preliminary clips was expected. However, the fact that their explanations referenced the failure of the system indicates that they interpreted the explanation generation task to be asking for an explanation of, "Why did the material spill on the floor?" as opposed to, "Why did the system work the way it did?". No participant in the non-systematic condition generated an explanation that had this focus. Instead, the participants in that condition interpreted the request for an explanation of the clip as something like, "Please explain the events within the clip," as opposed to, "Please explain what the experimenter is currently doing." This differentiation in explanatory focus is attributable to the participants in the different conditions selecting different contrast classes. Even though every participant saw the same final clip and entertained the same request for an explanation, they focused on different topics to explain. The participants in the systematic condition saw a final event that did not include a completion of the system unlike the events the participants had seen in the preliminary clips. The participants in the non-systematic condition saw another variation of events they had seen previously in the preliminary clips. The relationship of the final clip to the preliminary clips within each condition affected what the participants perceived as the explanandum. In the systematic condition, the explanandum was seen to be the failure of the material to go back into the oddly-shaped object. In the non-systematic condition, the explanandum was seen to be the series of events that led to the release of the material. This illustrates how the selection of the contrast class was driven by the context established by the preliminary clips.

Second, the content of the explanations also varied between the conditions. Most of the participants in the non-systematic condition used descriptions of the events and the causal relations between them, e.g. "the ball compressed and pushed against the rod causing the rope to pull…" They stayed very focused on the actual clip and what the saw happening in it. In contrast, the participants in the systematic condition made reference to other systems, most often biological systems. A number of the explanations from the systematic condition specifically reference "food," and several include the notion that the "beast" (or "plant-thing" or "organism") did not repeat the action in the final clip because it did not get the "food" it needed to continue. This type of content reflects the activation of general semantic knowledge, knowledge of living things and food and the consequences of not getting food. The activation of this knowledge seems to have been due to the similarity of the events in the systematic clips to a known functional system. It is important to note that access to this particular semantic knowledge was not itself responsible for the functional explanation. Five of the participants in the systematic condition provided functional explanations that did not reference a biological system. Several used a different source of semantic knowledge, referencing instead machines, and two provided generic functional explanations that did not reference any particular domain. We had intentionally designed the clips to be novel in the sense that we did not want the functional explanations to rely on an analogy to a known functional system. So, in each case, the functional explanations rely on the appreciation of the system, but the background knowledge affected how the participants chose to talk about that system. This indicates the recruitment of semantic knowledge during the explanatory process when relevance relations are selected.

These two patterns of results taken together indicate that the effect of background knowledge on explanatory processes is not as simple as we may have initially proposed. The results of this experiment suggest that different sources of knowledge are accessed depending on the stage of the explanatory process. When selecting the contrast class, information related to the immediate context is important to determine what is to be explained. When determining the relevance relations, semantic knowledge is important. However, as this is an initial inquiry, it is an open question as to whether semantic knowledge can affect the selection of the contrast class as well, and how knowledge of the immediate context might affect the relevance relations.

We propose that the background knowledge is critical to the explanatory process because the generation of an explanation relies on the development of an internal model of the explanandum. When prior experiences focus on a series of causal events with no larger system readily available, the participant constructs an internal model that emphasizes the causal relationships of the events. When the prior experiences instead capture the same events within a completed system, the system becomes an important aspect of the internal model and the subsequent explanations reflect that structure. The idea that these sorts of models underlie explanation is not new: Keil (2006) and Chaigneau, Barsalou, and Sloman (2004) have recently proposed similar ideas. However, by using the framework provided by the pragmatic theory, we can be more specific as to when particular information is incorporated and how it shapes the explanatory model that is constructed.

It appears that these models operate with a simplicity principle like that proposed by Lombrozo (2007). For the participants in the systematic condition, their explanatory model tended to be couched within the idea of a system, whether biological or machine-like. By invoking this system, they could generate an explanation for the final clip without going into the details of each action and event within the clip. Because of their appreciation of the larger structure of the system, the functional explanation became the simplest explanation. For the participants in the non-systematic condition, there is no a priori reason why they could not have made the inference that the oddly-shaped object would have a use for the material it so diligently collected during each clip, leading to an appreciation of a system similar to that of the participants in the systematic condition. However, it was simpler for those in the non-systematic condition to recount the causal chain that led to the final event and not make the additional inferences necessary to establish the presence of a system. In both cases, they could have made the other type of explanation, but it would have required more effort.

Despite the differences the conditions showed in the explanation generation task, they had a similar pattern of response in the explanation selection task. There are two questions here – why was there a difference between the two tasks and why the lack of a difference between the conditions in the explanation selection task? With regards to the first question, we argue that the two tasks present different challenges for the participants. In the explanation generation task, as we've explained prior, the participant has to identify the contrast class and then establish relevance relations once the topic of the explanation has been determined. We have argued that the two conditions showed a different pattern of responses because the prior experiences constrained these processes. However, during the explanation selection task, the participant is given the explanations to evaluate, bypassing the need to establish a contrast class or determine the relevance relations. Instead, the participants had multiple explanations specified and they simply had to determine the relative merit of each. With regards to the second question, it is possible that the causal explanation was selected as the best explanation by the majority of participants regardless of condition because of some quality of the particular explanation used (e.g. it was longer than the alternatives) or because people tend to like causal explanations best (e.g. Keil, 2006). Another possibility is that the participants in the systematic condition abandoned their functional stance because the underlying explanatory model they had developed was tentative and open to contamination from the explanations being evaluated during the selection task. Since the information available from the explanations was the same for all participants, this could have attenuated the differences between the conditions. Even though we did not fully articulate this idea ahead of collecting the data, it was the primary reason we had all participants do the explanation generation task before the explanation selection task. In prior studies where these types of evaluation tasks have been used successfully, participants were assessing less abstract explanatory models concerned with explananda like mops and echoes in caves, and the task was to compare across the fully specified alternative explanations. Our study had a different focus since we were primarily interested in the generation of the explanations.

In conclusion, we started this project with the assumption that background knowledge plays a critical role in explanation. By adopting the framework of the pragmatic theory for our inquiry, we were able to make some interesting and useful observations about how background knowledge is implicated in the stages of the explanatory process. The methodology we adopted in this study offers a rich opportunity for further study of explanation.

Acknowledgments

This study was supported by a Denison University Research Foundation grant. We thank Tania Lombrozo, David Landy, and the Ross-lab group at UIUC for sharing some thoughts on this project. Jessie Birdwhistell was invaluable in collecting the data for this study.

References

Bradner, A. (2005). *The End of Explanation.* Unpublished doctoral dissertation, Northwestern University, Evanston.

Chaigneau, S. E., Barsalou, L. W., & Sloman, S. A. (2004). Assessing the causal structure of function. *Journal of Experimental Psychology: General, 133,* 601-625.

Cummins, R. (1975). Functional analysis. *The Journal of Philosophy, 72,* 741-765.

Glymour, C. (1998). Learning causes: Psychological explanations of causal explanations. *Minds and Machines, 8,* 39-60.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57,* 227-254.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science, 48,* 507-531.

Leake, D. B. (1991). Goal-based explanation evaluation. *Cognitive Science, 15,* 509-545.

Lombrozo, T. (2006). The structure and function of explanations. *TRENDS in Cognitive Science, 10,* 464-470.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology,55*, 464-470.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99,* 167-204.

Salmon, W. C. (1988). Statistical explanation and causality. In J. C. Pitt (Ed.), *Theories of explanation*. New York, NY: Oxford.

Shanks, D. R., Holyoak, K. J., & Medin, D. L. (Eds.). (1996). *Causal learning.* San Diego, CA: Acad. Press.

Van Fraassen, B. C. (1988). The pragmatic theory of explanation. In J. C. Pitt (Ed.), *Theories of explanation.* New York, NY: Oxford.

Wright, L. (1973). Functions. *Philosophical Review, 82,* 139–168.

Appendix

Explanations Used in the Selection Task

Cause: "The motion of the odd-shaped thing initiated a sequence of events that ultimately resulted in yellow stuff coming out of the tall beaker."

Function: "The odd-shaped thing pushed a platform in order to get some yellow stuff out of the tall beaker."

Description: "The balloon was blue. The lever was red. The stuff in the beaker was yellow."

Wrong: "The beaker shattered after the lever hit, the platform melted, and purple stuff spilled over the balloon."

Example Explanations Generated by Participants

Systematic condition:

"In the previous clip, the plant-like thing that was using the ball and the board to feed itself, didn't utilize gravity like it had in the previous two, and because of that it couldn't get food and couldn't replenish its energy source which means it couldn't replenish itself with energy and because of that it didn't have the energy to push the ball again and continue to receive food." (functional explanation)

"The thing did not exert enough energy to move the ball in order to connect to the other cables so that it did not get to the food, or whatever, and it did not have the energy to repeat the action." (functional explanation)

"From what I saw, a large colorful blob pushed the rectangle into the blue ball. Because the ball couldn't, well its space was limited so it couldn't expand to the right, so it expanded up and down into an oval which bumped into… I think it was a bar, a red one. Well the bar went up – well the left side of the bar went up, so obviously the right side went down and pulled on a string that was connected to a pulley that pulled up on… that triggered the object on the right to release some yellow liquid." (causal explanation)

Non-Systematic condition:

"The creature pushes a block so the ball flexes, and the ball flexing pushes a lever which pulls a string, and this causes another block to pull up… um… which releases the fluid outside of the ball." (causal explanation)

"I think that the object that is creating the initial force is pushing down on a block of some sort which pulls a lever or pulley which then makes some kind of liquid come out." (causal explanation)

"I think the clip was about some alien animal with a trunk pressing the lever to get some lemonade out of like the dispenser then it's gonna slurp it up with its trunk." (functional explanation)