

Children and Pragmatic Implicatures: A Test of the Pragmatic Tolerance Hypothesis with Different Tasks

Katrijn Pipijn (Katrijn.Pipijn@Ppw.Kuleuven.Be)

Laboratory of Experimental Psychology, Tiensestraat 102 Bus 3711
3000 LEUVEN, Belgium

Walter Schaeken (Walter.Schaeken@Ppw.Kuleuven.Be)

Laboratory of Experimental Psychology, Tiensestraat 102 Bus 3711
3000 LEUVEN, Belgium

Abstract

The pragmatic tolerance hypothesis (Katsos & Smith, 2010) was originated to explain the difference between children and adults concerning scalar implicatures. They introduced the use of a Likert-scale to test this hypothesis. We conducted a study with a within subjects design in which we compare children's binary and scalar responses to the same underinformative sentences. We also used two separate tasks to look at the effects of task difficulty on performance. The results show that the more difficult task, Euler circles, lead to less pragmatic responses compared to the easier task, drawings. Confirming the study by Katsos and Smith (2010; see also Katsos & Bishop, 2011) children choose the middle options on the scale more when they are confronted with underinformative sentences and they choose more extreme options for the control sentences. The comparison with the binary responses however, reveal that the link between the two measuring methods is not as straight forward as we would think.

Keywords: Scalar Implicatures; underinformative sentences; children; scalar responses pragmatic tolerance.

Introduction

Communication is not always as straightforward as one might think. In 1989 Grice published his work on the cooperative principle that was meant to explain how our human interaction can be described. The cooperative principle expects a person to interact in a way that furthers the purpose of the conversation and indicates that a second person expects the first person to do so. The cooperative principle allows for implicatures to be used. When a person uses an implicature, the meaning of what that person says is not explicitly communicated, but can nonetheless be derived from what he says. The utterance is under-informative, more information could have been given but has not. For example when a wife asks her husband whether he'll be home for supper, and the husband answers that he has a meeting that will run late that day, then the husband is using an implicature. His wife will not expect him for dinner. One can assume that she accepts the meeting running late will be the reason, or at least a possible reason, that the husband will not be present at dinner. Nevertheless it is still possible that the husband will appear for dinner, for the implicature is cancellable. It is possible that the husband just meant he would be a little late for dinner, still he would not have lied in his earlier utterance.

One specific form of implicatures are scalar implicatures, which we will focus on in this paper. As the name implies, scalar implicatures consist of words that can be situated on a scale, known as Horn scales (see Horn, 1984). These words range from less informative to more informative, for example a scale containing words like *<none>*, *<some>* and *<all>*. Each word further on the scale contains more elements of a group. When a speaker uses a certain less informative word in an utterance, it is implicated that the more informative word is not applicable. When a person uses the word 'some', the word 'all' would not be appropriate. It is considered a mutual understanding between speaker and recipient that the speaker would have used the more informative word if it were suitable. Nevertheless he deliberately chose to use the less informative word on the scale therefore the more informative is not suitable. For example when the prime minister says 'Some banks are collapsing due to the financial crisis', a citizen can assume that 'not all' banks are collapsing due to this crisis, for the expression of 'some' implicates 'not all'. The citizen presumes that the prime minister would have said 'All banks are collapsing due to the financial crisis' if this were the case. If a few months later the prime minister makes the announcement 'All the banks have collapsed due to the financial crisis', this would not be a withdrawal of his earlier statement. Specific to implicatures is that they are cancellable in only one direction. When a speaker uses the weaker term 'some', it can later be easily corrected to 'all'. Yet when a speaker initially uses the stronger term 'all', it is not possible to change it to 'some' later on. At least not without admitting one was erroneous the first time. The stronger term 'all' entails the weaker term 'some' but not vice versa.

When a speaker uses the word 'some' in an utterance, there are two different ways to interpret this weak scalar term. The first way is the pragmatic way that was described above. A recipient might produce a scalar implicature and assume that the speaker meant 'some and not all' with the statement. Yet another way of interpreting the word 'some' is a purely explicit logical interpretation. The explicit meaning of the word 'some' is 'at least one and possibly all'. Both interpretation of the word are equally correct and it is the choice of the recipient on how he will interpret it.

Further in this article, we will refer to scalar implicatures as underinformative items or sentences.

We already know from different studies that children and adults interpret underinformative sentences in alternative ways. Noveck (2001) argues that a weak scalar term is understood in its explicit meaning first and will appear first in human development. Only later on the more complex pragmatic meaning will be incorporated. This argument is clearly demonstrated by the results of Noveck's study (2001). He found how children of 7-8 years old and 10-11 years old have acceptance rates of 89% and 85% for sentences that are logically true but pragmatically infelicitous. Adults on the other hand, accept these sentences in only 41% of the cases. This clearly demonstrated how for children the pragmatic meaning of these sentences is not incorporated. While for adults these pragmatic meanings are fully incorporated and are used as the principal criteria to accept or reject sentences.

The results also show how these differences between children and adults cannot be explained by the children's limited understanding of words like 'some' and 'all'. For all the different utterances that do not hold a conflict between the logical and the pragmatic meaning, the answering patterns of children and adults are very alike. The reason for the discrepancy between children and adults is not entirely clear. Noveck explains this by the posterior development of the pragmatic understanding of underinformative sentences. The processing of the pragmatic meaning of underinformative sentences is also cognitively much more demanding than the processing of the logical meaning (De Neys & Schaeken, 2007). Because of this, the pragmatic interpretation is harder to incorporate for children. Another factor that contributes to this is the nature of the task.

Pousoulous et al. (2007) reported experiments in which they changed the nature of the task from verbal judgments to action-based judgments. Using small boxes that contained tokens, participants were asked to alter the setting of the tokens to match a statement. They were also allowed to leave a setting as it was. Within the experimental design, children's performance on producing implicatures was much higher than in experiments with verbal judgments. This increased implicature production was found for all ages (4-, 5-, and 7-year-olds as well as adults). Still, the developmental effect was present. These experiments show how the understanding of implicatures can be facilitated in young children by changing task features. Other studies have also showed how changing task features can facilitate children's performance (Guasti et al., 2005; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004).

Katsos and Smith (2010) did research on underinformative sentences in children and adults. They raised the pragmatic tolerance hypothesis to explain for differences between children and adults as well as differences between adults. The starting point of this hypotheses is that there are different degrees of violations. Several violations can lie within an utterance yet not every violation is equally grave. Participants can and will reject

utterances that are a grave violation of the logical truth. Yet they might accept or reject an utterance that only holds a violation of informativeness and thus is an infringement of the cooperative principle. There is no implicit rule on how to deal with pragmatically infelicitous utterances. The threshold of what is and what is not acceptable is individual for each person and is called pragmatic tolerance by Katsos and Smith (2010).

An obvious way to test this hypothesis was adopted by Katsos and Smith (2011, also see Katsos and Bishop (2011) and Katsos et al (2011)). Katsos and Smith (2010) introduced the use of a Likert scale to the research on underinformative sentences. A Likert scale is a bipolar psychometric scale on which a participant can indicate to what extend he agrees or disagrees with a certain statement. Katsos and Bishop (2011) made their participants indicate how much they agreed with utterances containing the words 'some' and 'all'. Both children and adults clearly rejected utterances that were inherently false and accepted utterances that had an optimal use of the words 'some' and 'all'. Interestingly, for the underinformative utterances, the answering patterns for children and adults were also very similar, as both groups chose the middle option on a 3-point Likert scale. This is in strong contrast with Noveck (2001) were the answering patterns for children and adults were much more distinct, notwithstanding the children in this study were older. Katsos and Smith (2011) explain this with the pragmatic tolerance principle. Children appear to be competent pragmatic comprehenders. They do sense the pragmatic violation when underinformative sentences are used. Yet due to their different tolerance levels, they do not experience this violation to be grave enough to be rejected. Therefore, when they are confronted with a two alternatives forced choice, they will not reject the violation while adults will.

In this paper, we want to explore these results more thoroughly and make three hypotheses. First of all, we will vary the task method. Pousoulous et al. (2007) and others taught us that the nature of the task is of great importance. We expect that when we use different tasks, we will be able to make children reason more or less pragmatic, depending on the task difficulty. We will apply different methods than those used in Katsos and Bishop (2011) and Katsos and Smith (2010). Earlier research on underinformative sentences used different methods than the current ones. For example Newstead (1989, 1995) used Euler circles in his research. This abstract testing method should be difficult for children and thus induce more logical reasoning. We also developed a more child-friendly method using drawings which should induce more pragmatic reasoning in children.

Our second hypothesis concerns pragmatic tolerance. It seems obvious that this theory should be examined with a within subjects design in which children are confronted with a Likert scale as well as with the two alternative forced choice paradigm. We expand the testing method used in Katsos and Bishop (2011). Participants will be confronted with each underinformative sentence twice, once with the

option of responding on a Likert scale or once with a two alternative forced choice. With this research we expect to replicate Katsos and Bishops (2011) findings, namely that children do seem to detect a conflict when they are confronted with underinformative sentences. We expect that this conflict detection will be hidden when confronted with a two alternative forced choice but will become clear when they are confronted with the Likert scale. We will use children around the age of eleven, congruent with Noveck (2001). According to this study we expect children of this age to be still much more logical than adults.

Finally, we will look at consistency in children's answers. We expect that children that answer logically or pragmatically with the scalar measuring method, will answer in the same direction with the two alternative forced choice measuring method.

Method

Twenty-two Dutch speaking children participated in this research (mean age 11,3 range 11-13).

The children received a pen and paper test. The test started with a cover-up story about a boy named Thomas. The children were told that Thomas was new in class and came from a foreign country. They were told he was still learning the Dutch language and the children were to indicate how precise his answers were. Children had to indicate their answers either by indicating right or wrong, or on a 5-point Likert scale. The ends of the Likert scale were illustrated with a happy smiley and a frowning smiley. On the scale, the children were to indicate how well they thought that the boy's answer was, going from completely wrong to completely right. They were also allowed to use the middle options when the answer was only a little right or wrong or evenly right and wrong.

Two different tests were used. Both tests had the same basic structure. We started each trial with a given situation. This situation was presented either by a figure or a drawing. Then the participants were given a statement about the situation. They were instructed to indicate how well the statement described the situation given above.

First was the Euler circles task. The circles for each figure were either completely overlapping, partially overlapping or completely disconnected. Each circle represented a group of blocks, for example 'red blocks', 'square blocks', which was written inside each circle. The participants received a statement about the blocks and had to judge how precise the statement described the circles setting. For an example of this, see Figure 1.

For the second task, we used a method which was more adapted to children, Drawings. For the given situation, the children were now shown a drawing of a real life setting, for example a few kids playing with a bow and arrows. Again the children had to judge a statement about the setting, e.g. 'Some arrows are shot in the bull's-eye'. Due to the more authentic stimuli, the task became much easier for children.

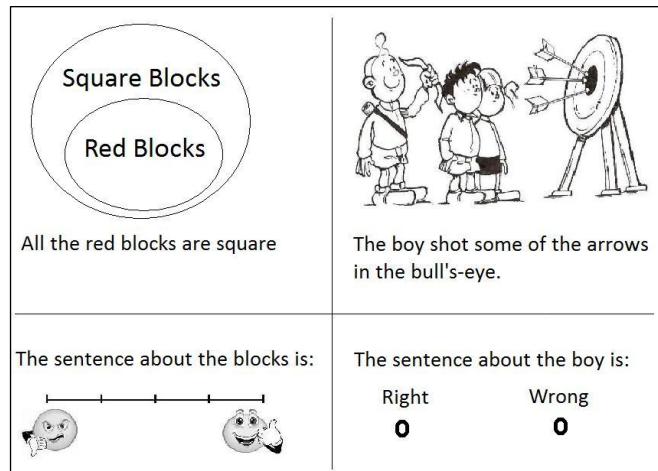


Figure 1: Example of Euler circles, drawings, scalar response option and binary response option.

Results

We inverted all scores of the logically false items. This way, high scores on the control items, for both logically false items and optimal items, indicate competent reasoning. We also inverted answers on the underinformative items. Because of this, the maximal score of five points is an extreme pragmatic answer and the minimal score of one is an extreme logical answer. Finally we converted the binary zero and one scores to one and five scores to make them comparable with the scalar responses.

For the control items we found very high average scores, 4.72 (.20) for binary responses and 4.56 (.34) for scalar responses. This means that the children understand the words 'some', 'all' and 'none' adequately. For the underinformative items, we found average scores of 3.93 (.89) for the binary responses and 3.16 (.98) for the scalar responses. For more detailed results, see Table 1.

Table 1: Mean ratings and standard error of the mean for Euler circles (EC) and Drawings (D)

	Binary	Scalar
EC – Control items	4.60 (.36)	4.46 (.36)
D – Control items	4.92 (.18)	4.65 (.41)
EC – Underinformative items	3.18 (1.51)	2.65 (1.05)
D – Underinformative items	4.70 (.57)	3.71 (1.11)

We ran a repeated measures design with three within factors with two levels each, namely measuring method, task and item type. We found three main effects. The two measuring methods levels, binary answers versus scalar answers, are significantly different from each other ($F(1,21) = 9.46$, $p < .01$). Binary responses are higher than scalar responses, as expected because binary responses only allow extreme answers. For the two tasks, Euler Circles seem to

be more difficult and lead to more logical answers than Drawings, $F(1,21) = 54.07$, $p < .00$. For the item types, control items versus underinformative items, children answer more extreme for control items and more varied for underinformative items, $F(1,21) = 54.72$, $p < .00$. We found two interaction effects. The interaction between measuring method and task was not significant but the other two interaction effects were, measuring method versus item ($F(1,21) = 4.63$, $p < .04$; see Figure 2) and task versus item ($F(1,21) = 21.62$, $p < .00$; see Figure 3). The three-way interaction was not significant.

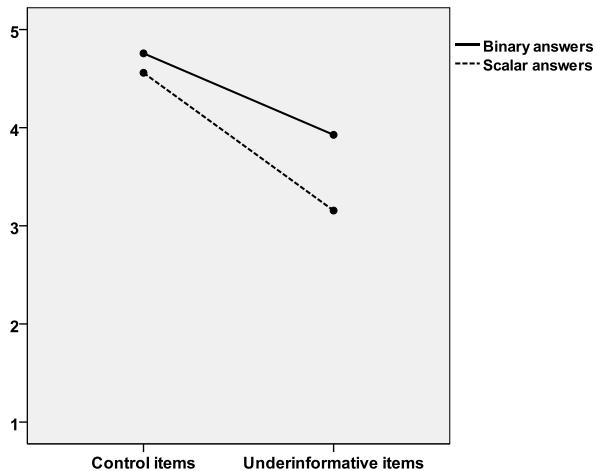


Figure 2: Interaction between measuring method and item type.

We calculated the difference between the control items and the underinformative items for each measuring method. A paired-samples t-test on these values was significant ($t(21) = 2.21$, $p < .04$). This means that the interaction between measuring method and item type is explained by a difference in size of the effect of measuring method on item type.

The main effect of task and its interaction with item, mean that the Euler Circles were more difficult, especially for the underinformative items and thus lead to more logical answers. To confirm this, we calculated the difference between the control items and underinformative items for each task and analyzed with a paired t-test, $t(21) = 4.65$, $p < .00$.

For the control items, 84% of the items were answered with an extreme answer of one or five on the scale. For the underinformative items, only 47% were answered with an extreme one or five. These two percentages were significantly different from each other ($t(21) = 5.22$, $p < .00$).

Finally we look at consistency of answers. We interpret being consistent between the two methods when a child gives an extreme answer of one or five on the scale and gives the equal binary response for the same item. For the control items, the children were fairly consistent between the two measuring methods. 80% of the children can be considered consistent under this rule. For the

underinformative sentences, children were much less consistent, only 33 % of them was consistent in their answers between the two methods. When we adopt a more flexible rule including also the two and four answers on the scale, which would also be acceptable, 87% and 57% of children can be considered consistent. For the underinformative items, 16% of the time the middle option of the scale was chosen.

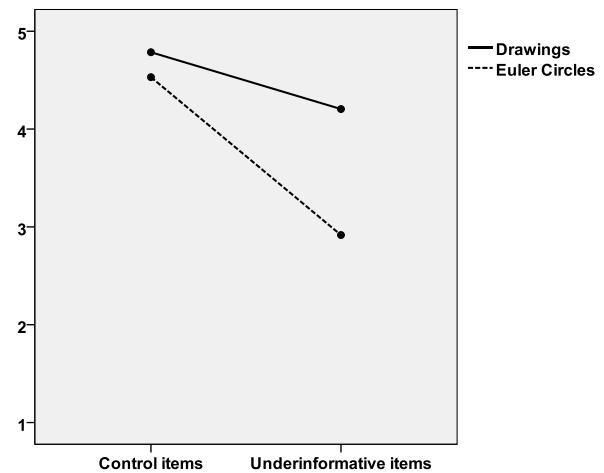


Figure 3: Interaction between task and item type.

Discussion

In this study we examined three hypotheses. First of all, we expected that children's performance will depend on the task difficulty. More precisely, we expected the Euler circles to be more difficult than the Drawings task and to lead towards less pragmatic answers for the underinformative items. Next we expected to replicate Katsos and Bishops (2011) findings, namely that children answer extremely pragmatic or logical when confronted with control items but more doubtful when confronted with underinformative items and a scale. Finally we expected children to be consistent in their answers on the two different measuring methods.

For the first hypothesis, we can find confirmation in the main effects of task and the interaction between task and item type. The Euler Circles task is clearly more difficult than the Drawings task. For the control items this difference is small but significant. For the underinformative items, this difference becomes even larger. For the more difficult task, the Euler circles, this leads to more logical answers. For the easier task, the Drawings, children become more pragmatic. There still remains a significant difference with the control items though. We hereby can confirm what Pouscoulous et al. and others made us expect. Task features can influence children's pragmatic reasoning on underinformative sentences. We noted earlier that we expect task difficulty to be the determining factor here. Yet we acknowledge that another factor may be at work as well. The Euler Circles task is believed to rely on logical reasoning skills. It might

be possible that the logical interpretation is triggered by the general logical characteristics of the task. In this case, not task difficulty but the logical nature of the task would be the determining factor. More in depth research on the matter seems necessary. The tasks used in this study were also very adapted to usage with children. More grammatical approaches to the material might lead to different conclusions. If the grammatical view of scalar implicatures (e.g. Chierchia, 2006; Fox, 2007) is correct, then in principle the implicature-computing operator could also be inserted in embedded positions, thus giving rise to embedded scalar implicatures. Chierchia, Fox and Spector (de twee papers) argue that an implicature-computing operator can indeed be inserted in embedded positions. It would be interesting to see how our conclusions and those of Katsos and Smith (2010) and Katsos and Bishop (2011) could be incorporated into this grammatical approach.

Secondly, we found a significant effect of measuring method and an interaction with item type. The difference between binary answers and scalar answers for the control items is significant. But the difference between the methods becomes much larger for the underinformative items. This confirms our hypothesis and replicates Katsos and Bishop (2011). When confronted with a scale, children do feel that there is a conflict between the pragmatic and the logical interpretation of underinformative sentences. They tend to choose the middle options of the scale more often (53%) than when confronted with control items (16%). This rules out the possibility that children are just unfamiliar with the use of scales. They are adequate in using scales and it is a deliberate action to choose the middle options for the underinformative items and the more extreme options for the control items. This confirms the pragmatic tolerance hypothesis in that children use the scale to express that they feel the conflict between the logical and the pragmatic interpretation.

We do however find a difference with common literature. The children in this study seem to be much more pragmatic than reports from other studies, especially with the binary responses. One explanation for this is probably the children's ages. Much research on this topic used younger children than the ones used in this study. It is self-evident that the slightly older children used in this study would perform more pragmatically and adult-like. Moreover, the current study was conducted in Dutch. Previous unpublished research on underinformative sentences with Dutch speaking children, revealed that these children are more pragmatic than their English-speaking (Katsos and Bishop, 2011) or French-speaking (Noveck, 2001) counterparts. Dutch speaking children seem to be more comparable to Spanish speaking children for example. In a study by Katsos et al. (2011), Spanish-speaking children rejected pragmatically false underinformative statements in 87% of the cases. It seems that the Dutch word 'sommige' is not the exact equal of the English word 'some'. This will probably contribute to the high rate of pragmatic answers in Dutch-speaking children.

Finally we examined consistency. These results seem to differentiate from the earlier found results. The children were not very consistent in their answers. Especially for the underinformative items, children were consistent in only 57% of the cases and 16% they chose the middle option. This still leaves 27% of the cases where children were not consistent. This percentage seems rather high to us and it interferes with the pragmatic tolerance theory. In roughly one fourth of the times, children's binary responses and their responses on the scale are not related. On top of that and in contrast to the study by Katsos and Smith (2010), we found much larger variances for both the control items and the underinformative items. This all suggests that the link between binary answers and scalar answers is not a direct link. For control items and underinformative items, up to 19% of the answers were cases in which the children gave an exact opposite to answer the binary items and the scalar items. We can hypothesize that in these cases children just made a simple error and that this wasn't intentional or due to a lack of understanding. But there is no way to be sure of this and it is in contrast with high overall levels of performance.

In conclusion, our study mainly confirms the pragmatic tolerance hypothesis but it also questions some aspects of it. It is clear to us that the pragmatic tolerance hypothesis and the relationship between binary and scalar answers on underinformative sentences is not as straightforward and that more thorough research on the matter is necessary.

Acknowledgments

This research was carried out with the financial support of the National Council for Scientific Research – Flanders, Belgium (FWO grant G.0634.09)

References

- De Neys, W., & Schaecken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, 54, 128-133.
- Chierchia, G. (2006). Broaden Your Views: implicatures of Domain Widening and the "Logicality" of Language. *Linguistic Inquiry*, 37, 535-590.
- Chierchia, G., Fox, D., & Spector, B.. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. To appear in Portner, P., Maienborn, C. & von Heusinger, K. (Eds.), *Handbook of Semantics*. New York, NY: Mouton de Gruyter.
- Chierchia, G., Fox, D., and Spector, B. (2009). Hurford's constraint and the theory of scalar implicatures: evidence for embedded implicatures. In Egré, P. & Magri G. (Eds.), *Presuppositions and implicatures. Proceedings of the MIT-Paris workshop*. MIT Working Papers in Linguistics 60.
- Fox, D. (2007). Free choice and the theory of scalar implicatures. In Sauerland, U. & Stateva, P. (eds.), *Presupposition and implicature in compositional*

semantics, 71–120. Hounds Mills, Basingstoke: Palgrave Macmillan.

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press, Cambridge MA.

Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, 20, 667-696.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference. In Schiffarin, D. (ed.), *Meaning, Form and Use in Context: Linguistic Applications. Proceedings of GURT '84*. Washington D.C.: Georgetown University Press.

Katsos, N., Andrés Roqueta, C., Estevan, R. A. C., & Cummins, C. (2010). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification? *Cognition*, 119, 43–57.

Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition*, 120, 67-81.

Katsos, N., & Smith, N., 2010. Pragmatic Tolerance or a speaker–comprehender asymmetry in the acquisition of informativeness? In Franich, K., Iserman, K.M., Keil, L.L. (Eds.), *Proceedings of the 34th Annual Boston Conference in Language Development*, Cascadilla Press, MA, USA.

Newstead, S. E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language*, 28, 78-91.

Newstead, S. E. (1995). Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language*, 34, 644-664.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78, 165-188.

Papafragou, A., & Musolino J. (2003). Scalar implicatures: experiments at the semantics/pragmatics interface. *Cognition*, 86, 253-282.

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition*, 12, 71-82.

Pouscoulous, N., Noveck, I., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, 14, 347-376.