Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 3, 687–696

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.3.687

Actions and Affordances in Syntactic Ambiguity Resolution

Craig G. Chambers
University of Calgary

Michael K. Tanenhaus
University of Rochester

James S. Magnuson
Columbia University

In 2 experiments, eye movements were monitored as participants followed instructions containing temporary syntactic ambiguities (e.g., "Pour the egg in the bowl over the flour"). The authors varied the affordances of task-relevant objects with respect to the action required by the instruction (e.g., whether 1 or both eggs in the visual workspace were in liquid form, allowing them to be poured). The number of candidate objects that could afford the action was found to determine whether listeners initially misinterpreted the ambiguous phrase ("in the bowl") as specifying a location. The findings indicate that syntactic decisions are guided by the listener's situation-specific evaluation of how to achieve the behavioral goal of an utterance.

As a sentence unfolds in time, the grammatical relationships among its constituents are often temporarily ambiguous. For example, the phrase italicized in (1) may indicate the location where an egg is being poured, or may specify which of several eggs is intended.

(1) The baker poured the egg *in the bowl*. . .

(2) a. . . .while stirring continuously.

b. . . .over the flour.

Although information that follows the ambiguity typically clarifies the intended grammatical relationship (e.g., 2a–2b), there is substantial evidence that listeners assign a provisional grammatical analysis to the constituent as soon as it is encountered. Two central questions in research on sentence processing focus on the kinds of information used to make this decision and how (or when) this information is used. In *modular* models, syntactic constraints, or a subset of syntactic constraints, play a privileged role in initially selecting a single structure, or in ranking alternative structures, on the basis of characteristics of the evolving constituent structure (Frazier, 1987; Frazier & Clifton, 1996). Other information sources that could in principle guide this decision (e.g., semantic and discourse-based knowledge) are not used until a later stage of processing where the original analysis is either confirmed or revised.

In contrast, *constraint-based* approaches maintain that frequency-weighted syntactic alternatives are continuously evaluated using a range of constraints, including nonlinguistic information sources (MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995). To date, two main classes of nonlinguistic constraints have been identified (Gibson & Pearlmutter, 1998; but cf. Binder, Duffy, & Rayner, 2001). The first class is plausibility. The difficulty arising from processing infrequent (dispreferred) structures is attenuated when the normally preferred analysis is implausible (Trueswell, Tanenhaus, & Garnsey, 1994). For example, on hearing "The evidence examined. . .", the normally dispreferred relative clause analysis (as in "the evidence examined by the lawyer was. . .") is more strongly considered because *evidence* cannot plausibly be the agent of the verb *examine* (also see Clifton et al., 2003).

The second nonlinguistic constraint is referential context. Many temporary ambiguities follow definite noun phrases (e.g., *the egg* in Example 1), which are used to refer to a uniquely identifiable referent. The amount of information necessary to identify a unique referent depends on the presence of other possible candidates in the referential context. This in turn may influence the grammatical role initially assigned to an ambiguous phrase (Crain & Steedman, 1985). For example, if several eggs are present, the listener may interpret the ambiguous phrase in (1) as a modifier because additional information is necessary to distinguish the intended egg. If only a single egg is present, modifying information would not be necessary and the listener would likely adopt a goal analysis where the phrase indicates the intended location. This prediction has been supported in a range of experiments showing that, in contexts with multiple potential referents, the confusion that typically arises with

Craig G. Chambers, Departments of Psychology and Linguistics, University of Calgary, Calgary, Canada; Michael K. Tanenhaus, Department of Brain and Cognitive Sciences, University of Rochester; James S. Magnuson, Department of Psychology, Columbia University.

This research was supported by grants from the National Institutes of Health (HD-27206), National Institute on Deafness and Other Communication Disorders (DC-005765), and Natural Sciences and Engineering Research Council (Canada; 69-6157). Special thanks go to Dana Subik for his invaluable assistance with the experiments. We also thank Keith Rayner and two anonymous reviewers for helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Craig G. Chambers, Department of Psychology, University of Calgary, 2500 University Drive, NW, Calgary, AB T2N 1N4 Canada. E-mail: craig@ucalgary.ca

the (dispreferred) modifier interpretation of a temporarily ambiguous phrase is reduced or eliminated (e.g., Altmann & Steedman, 1988; Britt, 1994; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Sekerina, Hill, & Logrip, 1999; but cf. Binder et al., 2001).

Nonlinguistic Constraints and Modularity

The most influential modular models of syntactic processing are typically serial parsing systems in which the initial phase of processing has access only to information about configurational syntax. However, neither the classic articulation of the modularity hypothesis (e.g., Fodor, 1983) nor some recent interpretations (e.g., Coltheart, 1999) place such stringent limitations on the kinds of information available to a modular linguistic processing system. Rather, the hallmark of modular processing is the inability of higher level domain-general factors to penetrate into the processing module. In the words of Fodor (1983)

> to show that that system is penetrable (hence informationally unencapsulated), you would have to show that its processes have access to information that is not specified at any of the levels of representation that the language input system computes; for example, that it has generalized access to what the hearer knows about the probable beliefs and intentions of his interlocutors. (p. 77)

According to the encapsulation view, a modular system could in fact involve the simultaneous use of various information sources functioning in either a bottom-up or top-down manner, so long as these information sources are computed at some level of representation within the language module. Thus, although effects of plausibility and referential context are problematic for certain classes of processing models, they are not necessarily inconsistent with this broader notion of modularity because they can be understood in terms of information that is intrinsic to a linguistic system. For example, mechanisms that simply keep track of co-occurrences among words and structures can simulate plausibility effects without invoking semantic or real-world knowledge (Burgess & Lund, 1997). Likewise, contemporary semantic theories propose that referents evoked in a discourse are enumerated in a representation that is an integral component in semantic interpretation (Heim, 1982; Kamp & Reyle, 1993). Referential effects in syntactic processing could therefore involve shallow computational mechanisms in which form-based presuppositions trigger mappings to entities in a semantic representation of discourse (e.g., Gordon & Hendrick, 1998). In light of these possibilities, evaluations of modularity in syntactic processing require manipulating informational constraints that are not arguably encoded in linguistic representations.

Domains of Interpretation in Real-Time Processing

Issues of modularity aside, additional motivation for understanding the interaction between linguistic and nonlinguistic constraints comes from recent situated and embodied approaches to language behavior (e.g., Barsalou, 1999; Glenberg & Robertson, 2000; Stanfield & Zwaan, 2001). Most relevant to the current study is work investigating how domains of referential interpretation are established as a sentence unfolds in time. In one study, Altmann and Kamide (1999) monitored eye movements during a sentence–picture verification task for sentences such as, "The boy will move/eat the cake." The results showed that listeners began to fixate the target object in the scene (e.g., a cake) on hearing the verb when the verb information was compatible only with the target object (e.g., *eat* but not *move*). This outcome is consistent with evidence that verbs activate typical event participants (e.g., McRae, Ferretti, & Amyote, 1997) but, more important, shows that listeners use predicate-based information to anticipate upcoming referents.

In related work, Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002) investigated whether real-world properties of objects constrain the referential domain used in the semantic evaluation of noun phrases. Eye movements were monitored as listeners followed instructions to move real objects (e.g., "Pick up the cube. Now put it inside the/a can"). The results showed that the semantics of *inside* and the action-relevant properties (i.e., *affordances*) of candidates jointly restricted the referential domain to only those candidates compatible with the denoted action (e.g., containers that were large enough to accommodate the cube). In addition, when the cube could fit inside only one of two visually available cans, listeners experienced difficulty interpreting indefinite (e.g., *a can*) but not definite (e.g., *the can*) noun phrases. The opposite pattern occurred when both cans could accommodate the cube. These outcomes suggest that linguistically relevant referential domains are continuously defined by evaluating sentence information against the properties of candidates in the contextual environment, including situation-specific affordances that are not encoded as part of the linguistic representation of a word or phrase. In the present study we extend these investigations by evaluating the consequences of action-defined referential domains for syntactic processing. Our goal is to investigate how the listener's evaluation of action-relevant referents influences the identification of grammatical relationships and, in doing so, to provide a test of the broader notion of modularity outlined earlier. In Experiment 1 we evaluate how initial syntactic decisions are affected by the number of referential candidates that can afford the action evoked by the verb in the unfolding sentence. In Experiment 2, the relevant affordances of candidates are manipulated by providing the listener with an instrument for performing the action.

Experiment 1

An example instruction from Experiment 1 is shown below. In (3a), the first prepositional phrase ("in the bowl") is temporarily ambiguous between the obligatory argument that specifies the intended location (goal) of the theme object (i.e., the egg) or a modifier that provides information about the theme object. The copular complementizer *that's* is used to create an unambiguous version of the instruction (3b).

(3) a. Pour the egg in the bowl over the flour.

b. Pour the egg that's in the bowl over the flour.

In a previous visual world study, Spivey et al. (2002) found that, on hearing ambiguous phrases such as that in (3a), listeners typically looked to a false goal location (e.g., an empty bowl in the display) when only one candidate referent for the theme object was present, indicating that a goal analysis of the phrase was initially adopted. This was typically followed by a fixation to the true goal (e.g., the flour) when the following prepositional phrase was heard. However, when two referential candidates were present in the display, fixations to the false goal location were reduced and were no more frequent than in conditions with unambiguous instructions, suggesting that a modifier analysis of the phrase was initially pursued.

We constructed two versions of the display corresponding to a given critical instruction (see Figure 1). Each version contained four object types: the target referent (top right, egg in bowl), the referential competitor (top left, egg in glass), a true goal (bottom right, flour), and a false goal (bottom left, empty bowl). What varied across the two versions of the display were the properties of the referential competitor. In the compatible competitor condition (Panel A), the competitor and the target are both in liquid form and, thus, are both compatible with the action denoted in the instruction (i.e., "pouring"). In the incompatible competitor condition (Panel B), the competitor is a solid egg and therefore cannot be poured. If syntactic decisions reflect an assessment of which candidates can afford the action evoked in the unfolding utterance, the listener should be "garden-pathed" (i.e., should adopt a goal analysis of the ambiguous phrase) in the incompatible competitor condition but not in the compatible competitor condition. This is because the referential domain in the incompatible condition should be narrowed to include only the pourable egg, thereby eliminating the need for modification. Alternatively, if syntactic decisions are informed only by information encoded in linguistic representations (e.g., the number of candidates meeting the description of the term *egg*), listeners should be garden-pathed in both conditions with ambiguous instructions. Following Spivey et al. (2002), we expected the grammatical role that is provisionally assigned to the ambiguous phrase to be reflected in fixations to the false goal object when this phrase is heard.

Method

Participants. Sixteen native speakers of English from the University of Rochester community received payment for participation.

Materials. We constructed eight experimental instructions similar to the examples shown in (2) above. Two versions of each instruction were prepared. In one version (e.g., Example 3a), the first prepositional phrase was temporarily ambiguous. In the other version (e.g., Example 3b), the complementizer *that's* explicitly signaled that the first prepositional phrase was a modifier. The corresponding visual displays consisted of real objects placed on stair-like shelves mounted on a table. On critical trials, each display contained the four item types described above. The target referent was always compatible with the event denoted by the main verb. The referential competitor was varied such that it was compatible with the verb event on 50% of trials and incompatible on remaining trials. The competitor manipulation (compatible–incompatible) was crossed with the two versions of the instruction (ambiguous–unambiguous), yielding four conditions. Four counterbalanced lists were constructed by varying the condition for each stimulus item across lists such that each list contained two stimulus items corresponding to a given condition, and each stimulus item appeared in a given condition only once in each list. The positions of the various objects in the display were counterbalanced across critical trials.

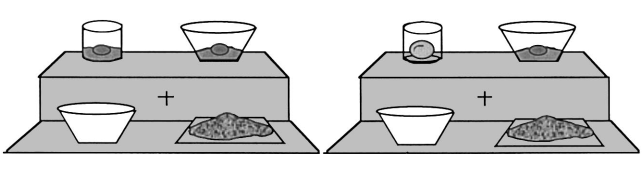

Figure 1. Example display, Experiment 1.

In addition to the critical stimuli, 32 filler instructions were prepared. Eight of the filler instructions followed critical instructions and referred to objects in the same display. The remaining 24 filler instructions were paired with 12 distinct filler displays. The filler displays and instructions were designed to prevent participants from recognizing particular contingencies and from adopting strategies on critical trials. For example, a number of filler displays resembled critical trials in that they contained two (or more) objects that could be denoted by the same noun and that contrasted with one another on the basis of a few functional attributes. However, these objects were either (a) never referred to, (b) not differentiated in terms of their contrasting attributes, and/or (c) their contrasting attributes were not relevant for selecting objects for the instructed action. In addition, a number of filler instructions contained a verb followed by a noun phrase and only a single prepositional phrase, or by a sequence of two prepositional phrases where the second modified the noun in the first (e.g., "put the spoon in [the cup on the saucer]").

The full set of instructions was recorded on a PC using a KAY speech acquisition and analysis system (Kay Elemetrics, Lincoln Park, NJ). The stimuli were recorded at 22050 kHz using a 16-bit sample size. The speaker took care to avoid producing prosodic cues that would disambiguate the ambiguous preposition phrase in the critical instructions. Two tokens of each critical instruction were recorded, and the most neutral token was selected for presentation with both versions of the visual display. The audio files for the individual instructions were then ordered and assigned to their respective lists. The files were played using a PsyScope program (Cohen, MacWhinney, Flatt, & Provost, 1993) running on a Macintosh G3 computer.

Procedure. Participants were tested individually. They were seated in front of the display table, which was adjusted to accommodate their height and reach. They were told that they would hear a sequence of instructions to manipulate the objects on the table top and that they should follow the instructions in a natural manner, including asking for clarification when necessary. They were then given several example instructions. After the examples, participants were fitted with a head-mounted eye-tracking device (Model 501, Applied Scientific Laboratories, Bedford, MA). The device consists of a lightweight eye camera and video scene camera attached to an adjustable headband. The eye camera provides an infrared image of the participant's left eye sampled at 60 Hz. Relative eye in-head position is calculated from the image by tracking the center of both the pupil and the first Purkinje corneal reflection. The video scene camera provides an image of the environment from the perspective of the participant. The scene image is displayed on a television monitor with superimposed cross-hairs that indicate the participant's point of fixation. The accuracy of the resulting eye-movement record is within 1° of visual angle across a range of ±20°.

A brief calibration procedure was conducted at the beginning of the experiment to map eye position coordinates onto corresponding scene image coordinates. A Hi8 VCR was used to record the image on the television monitor along with the instructions, which were presented over speakers located on each side of the display.

Two practice trials preceded the experiment to reconfirm that participants understood the procedure. After each instruction, one experimenter removed the items from the display and a second experimenter prepared the display for the next trial. The first experimenter cued each instruction by pressing a key on a keyboard located behind the participant.

The accuracy of the eye-movement record was monitored throughout the experiment, and minor adjustments were made between trials when necessary. The entire session lasted approximately 40 min.

Results and Discussion

Data were analyzed using frame-by-frame playback of the videotapes with the video and audio channels synchronized. The onsets and offsets of the spoken words in the critical instructions were recorded along with the point of fixation for each video frame. Figure 2 shows the proportion of fixations to the false goal object (e.g., the empty bowl in Figure 1). The data are shown for a 2,500-ms window following the onset of the noun in the first prepositional phrase (e.g., the onset of *bowl* in Example 3). The endpoint of this window corresponds to when participants had typically initiated a hand movement toward the target. The vertical lines are aligned with the average onsets or offsets of selected speech landmarks, as indicated.

The data show that false goal fixations in the incompatible competitor condition with ambiguous instructions increased toward the offset of the noun and remained at a fairly stable level throughout the region. In contrast, false goal fixations in the other conditions were initiated at a slower rate and were less sustained overall, showing very few fixations after 300 ms following the offset of the final noun. Note that some fixations to the false goal object are expected even when a modifier analysis is pursued because the listener may anticipate that this object will be the subsequently mentioned goal, as in, "pour the egg in the bowl in the other bowl/in the bowl on the right."

To provide a statistical analysis of the results, the percentage of time spent fixating the false goal was calculated within the 2,500-ms window. The means for each condition are shown in Figure 3 (additional measures are reported in Appendixes A and B). This measure captures both the number and the duration of fixations to the false goal object (fixations that began prior to the left boundary of the window are excluded). These values were submitted to an analysis of variance (ANOVA) with competitor compatibility and instruction type as within-participant factors. Because each array of display objects occurred in each of the conditions, the only analyses we report are those treating participants as a random variable (Raaijmakers, Schrijnemakers, & Gremmen, 1999). The analysis revealed a main effect of competitor compatibility, $F(1, 12) = 5.63$, $p < .05$, reflecting the increased time spent fixating the false goal when the referential competitor was not compatible with the required action. There was also a marginal effect of instruction type, $F(1, 12) = 4.27$, $p = .06$, reflecting the greater percentage of time spent fixating the false goal in the conditions with ambiguous instructions. These effects were qualified by a significant interaction of competitor compatibility and instruction type, $F(1, 12) = 8.68$, $p < .05$. Planned comparisons indicated that, with ambiguous instructions, the time spent fixating the false goal was greater when the competitor was not compatible with the required action, $F(1, 12) = 16.81$, $p < .01$. In addition, the time spent fixating the false goal in the incompatible competitor condition was longer when ambiguous instructions were used compared with when unambiguous instructions were used, $F(1, 12) = 15.71$, $p < .01$, and compared with when unambiguous instructions occurred with a compatible referential competitor, $F(1, 12) = 15.19$, $p < .01$. No other contrasts were reliable (Fs < 1).

These results suggest that visual contexts with two (compatible) referents completely eliminated the usual preference for a goal interpretation of the temporarily ambiguous phrase, replicating previous studies (Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al., 1999). Crucially, however, the goal misanalysis was reinstated when only one of the two potential referents was compatible with the action required by the instruction. This outcome provides support for the idea that properties of candidate referents are assessed in terms of their compatibility with the goal of the unfolding utterance, in turn affecting syntactic decisions.

The results clearly demonstrate that a nonlinguistic domain restriction can influence the earliest moments of syntactic ambiguity resolution. Moreover, the relevant domain cannot be established without canvassing the situation-specific properties of objects. Thus, the effect is clearly not semantic. However, it is possible to argue that action-based affordances were able to influence syntactic ambiguity resolution only because evaluation of the relevant attribute was triggered by information encoded in the lexical representation of the verb. For example, the lexical representation of the verb *pour* might specify that its theme must be liquid. In fact, the proposal that certain syntactically relevant

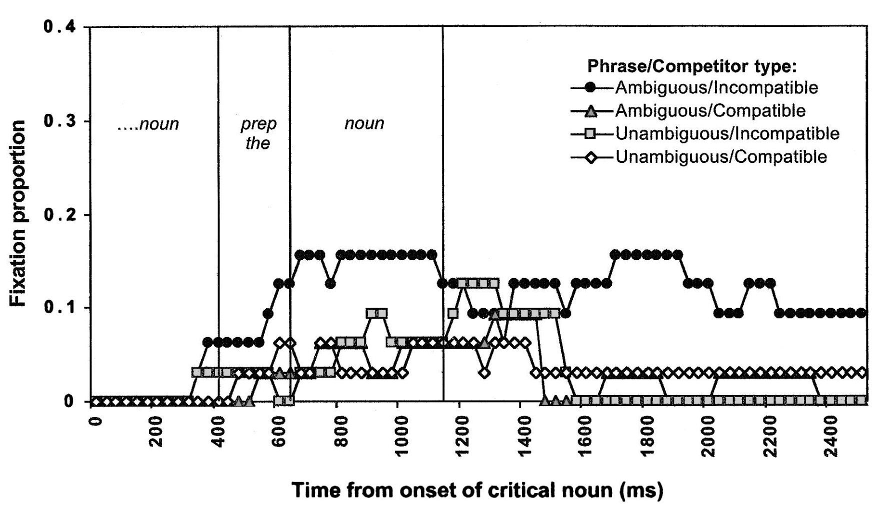

Figure 2. Proportion of fixations to false goal object over time, Experiment 1.

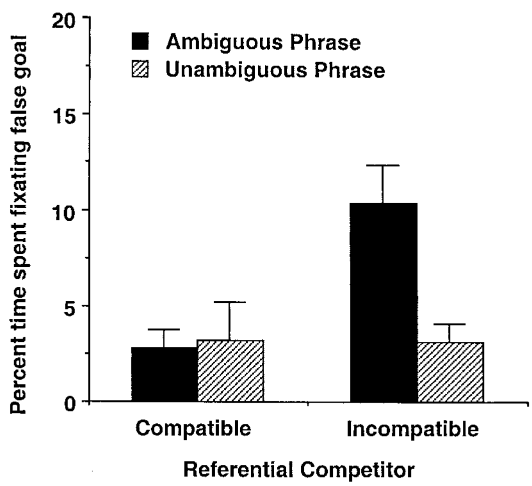

Figure 3. Percentage time spent fixating false goal object in specified time window, Experiment 1.

semantic features, so-called "selectional restrictions", are incorporated into the representations of verbs dates back to Chomsky (1965). To rule out a selectional restriction account, then, it is necessary to show that syntactic ambiguity resolution is influenced by action-based affordances when the relevant dimension is not evoked by information that is arguably incorporated into linguistic representations. In Experiment 2 we created these conditions by using the neutral verb *put* and by introducing constraints on possible actions by providing the participant with an instrument for performing the action.

Experiment 2

An example instruction for Experiment 2 is shown in (4). Figure 4 depicts the corresponding display, which contains two candidate whistle referents, one of which (the target referent) is attached to a loop of string. Importantly, the semantics of the verb *put* cannot determine which whistle must be used to complete the action required by the instruction.

(4) Put the whistle (that's) on the folder in the box.

In this experiment, the compatibility of the referential competitor (e.g., the whistle with no string) was manipulated by varying whether participants were provided with an instrument object to perform the described action. For example, before participants heard the instruction in (4), they might be given a small hook. Critically, this hook could not be used to pick up the competitor whistle, which did not have a string attached. If inferences about the intended action are generated on hearing, "Put the. . .", the presence of the instrument could result in the elimination of the competitor from the referential domain. If so, results should be comparable to Experiment 1. In particular, participants should misinterpret "on the folder" as the goal only when ambiguous instructions are used and an instrument is provided. If, however, the effects observed in Experiment 1 are driven primarily by linguistic information encoded in the main verb, and not judgments about possible actions, a modifier analysis should occur regardless of whether an instrument is supplied beforehand.

Method

Participants. Twenty-four native speakers of English from the University of Rochester community were recruited from posted notices and paid for their participation. None had participated in Experiment 1.

Materials. Twelve experimental instructions were constructed using the same sentence frames and display types used in Experiment 1. The presence or absence of an instrument was crossed with the two versions of the instruction (ambiguous–unambiguous) to create four conditions. We avoided semantic associations between the instrument name and the name of the intended target referent, using pairs such as hook–whistle, tweezers–candle, and chopsticks–matches. Four counterbalanced lists were constructed by cycling stimulus items across lists such that each list contained three different stimulus items in a given condition.

In addition to the critical stimuli, 40 filler instructions were prepared. Twelve of these followed the critical instructions. The remaining 28 filler instructions were paired with 14 distinct filler displays. As before, the filler displays and instructions were designed to prevent participants from recognizing particular contingencies on critical trials. In this experiment it was important to ensure that participants could not infer the identity of the target referent when they were provided with an instrument. To prevent this strategy, instruments were provided for 7 filler instructions (note that only six experimental trials had accompanying instruments). On 4 of these filler trials, the instruction required the participant simply to place the instrument in a specified location, rather than use it to perform an action on objects in the display. On the remaining 3 instrument filler trials, the instruments were used to perform actions on display objects, but not objects that would be typically associated with the instrument, nor were the actions typical of the instrument. Thus, the presence of an instrument on a given trial was neither a reliable predictor of whether the instrument would be used to manipulate a given object or whether the instrument would even be used to pick up an object.

The remaining filler instructions and displays were designed to balance other contingencies in the linguistic and visual stimuli and resembled those described in the previous experiment. Overall, 38% of fillers contained verbs other than *put* (e.g., *rotate, push, touch*) to reduce the likelihood that participants might preferentially attend to how display objects might be picked up. The linguistic stimuli were recorded and tokens were selected according to the method described in Experiment 1.

Procedure. The procedure was the same as Experiment 1 except that participants were told they would sometimes be given an object to use in carrying out the following instruction. An example was provided during the explanation of the task, and one of the practice trials also included an instrument. On instrument trials, the experimenter always handed the instrument to the participant without naming it.

Results and Discussion

Because of a minor error in the pairing of speech files to conditions, data from 8 trials were excluded from analysis (2.8% of total data). Figure 5 shows the proportion of fixations to the false goal object for a 2,500-ms window following the onset of the noun in the first prepositional phrase (e.g., the onset of *folder* in

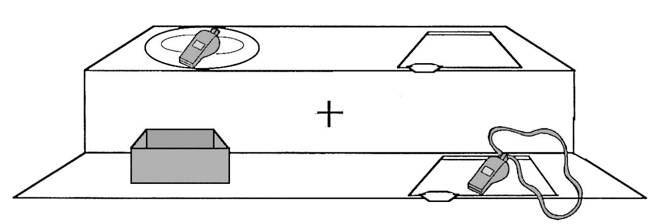

Figure 4. Example display, Experiment 2.

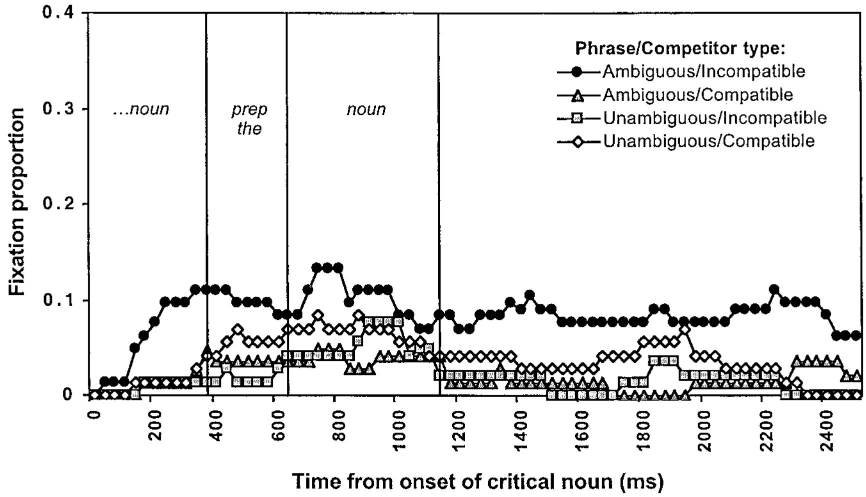

Figure 5. Proportion of fixations to false goal object over time, Experiment 2.

Example 4). The pattern of fixations was comparable with that observed in Experiment 1. Specifically, the largest proportion of fixations to the false goal occurred when ambiguous instructions were used and when the referential competitor was not compatible with the instructed action. Fixations to the false goal also began earlier in this condition. The results for the remaining three conditions shared the same general profile of comparatively fewer and slower fixations to the false goal.

Figure 6 shows the percentage of time spent fixating the false goal object within the defined temporal window for each experimental condition. Again the data pattern is similar to Experiment 1: The percentage of time spent fixating the false goal was largest in the condition where ambiguous instructions were used and the competitor could not afford the evoked action. The remaining conditions all showed a relatively smaller amount of time spent fixating the false goal. The data were submitted to a 2 × 2 ANOVA with competitor type (incompatible–compatible with required action) and instruction (ambiguous–unambiguous) as within-participant factors. The analysis revealed a significant interaction, $F(1, 20) = 11.00$, $p < .01$, and no significant main effects. Planned comparisons indicated that the percentage of time spent fixating the false goal in the incompatible competitor condition with ambiguous instructions was reliably larger than in the corresponding condition with a compatible competitor, $F(1, 20) = 14.54$, $p < .01$; the corresponding condition with unambiguous instructions, $F(1, 20) = 14.90$, $p < .01$; and the condition with unambiguous instructions and a compatible competitor, $F(1, 20) = 8.90$, $p < .01$. No other comparisons reached significance (Fs < 1).

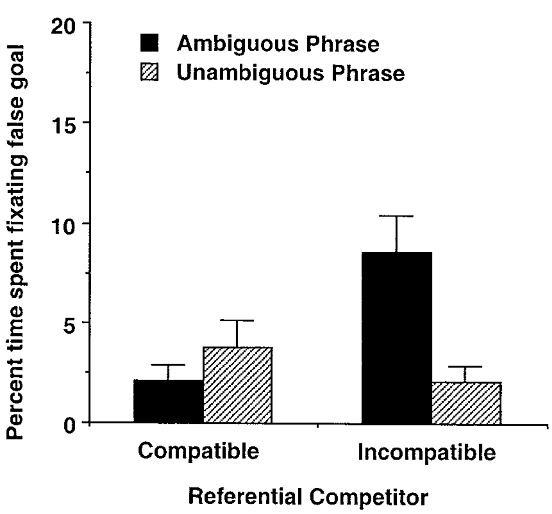

Figure 6. Percentage time spent fixating false goal object in specified time window, Experiment 2.

We also analyzed Experiments 1 and 2 together in a mixed-model ANOVA using competitor type (compatible–incompatible with action) and instruction type (ambiguous–unambiguous) as within-participant factors and "constraint source" (as manipulated across experiments: verb–instrument object) as a between-participants factor. The results revealed significant main effects of instruction type, $F(1, 38) = 6.15$, $p < .05$, and competitor type, $F(1, 38) = 5.78$, $p < .05$. These effects were qualified by an Instruction Type × Competitor Type interaction, $F(1, 38) = 17.01$, $p < .001$. The main effect of constraint source was not significant, nor did it enter into any significant interactions (Fs $<$ 1). Contrasts collapsing across constraint source indicated that more time was spent fixating the false goal in the ambiguous instructions when the referential competitor was incompatible compared with the corresponding compatible competitor condition, $F(1, 38) = 26.64$, $p < .001$, and compared with the unambiguous conditions with either incompatible or compatible competitors, $F(1, 38) = 25.68$, $p < .001$, and $F(1, 38) = 19.34$, $p < .001$, respectively. No other comparisons approached significance (Fs < 1).

Together, the two experiments provide evidence that the syntactic role assigned to a temporarily ambiguous phrase varies according to the number of referents that can afford the action evoked by the unfolding instruction. The same results hold regardless of whether the constraints on possible actions are introduced linguistically by a verb (Experiment 1) or nonlinguistically by the presence of a task-relevant instrument (Experiment 2). Therefore, the referential domain for initial syntactic decisions is determined by the listener's consideration of how to execute the action that constitutes the behavioral goal of the instruction, rather than by information sources that can be isolated within the linguistic system.

General Discussion

In this study we investigated whether syntactic decisions are informed by the listener's assessment of referents that are compatible with the action evoked in an unfolding utterance. Our goal was to investigate how perceptual and action-based knowledge is used in the course of language understanding and to evaluate whether this information constrains the earliest moments of syntactic processing.

In Experiment 1 we showed that listeners restricted attention to candidates that could afford the action evoked in an unfolding instruction. This in turn determined the syntactic role assigned to an ambiguous constituent. For example, as the instruction, "Pour the egg in the bowl over the flour", unfolded, listeners expected *in the bowl* to be a modifier when the scene contained two eggs in liquid form, reflecting the need for additional information to individuate the intended referent. However, when one of the two eggs was in solid form, listeners expected the phrase to specify the goal (i.e., intended location) for the single pourable egg.

In Experiment 2 we evaluated the possibility that the restricted referential domain is defined by selectional restrictions encoded in the main verb (e.g., *pour*), perhaps requiring only minimal access to nonlinguistic information. Critical instructions all contained the neutral verb *put* and referential candidacy was manipulated by providing the listener with an instrument object. Most generally, the results suggest that the referential domain relevant to syntactic processing is defined by the listener's assessment of possible actions in the given environment. Below we discuss the implications of the results for understanding how interpretive domains are established during processing and for the proposal that syntactic processing is modular.

Domains of Interpretation in Language Processing

The current results add to the growing body of evidence suggesting that the comprehender's knowledge of possible actions in a given situation is used to circumscribe linguistically relevant context. In previous work (Chambers et al., 2002), we found that predicate information and the affordances of referential candidates jointly define the domain used to interpret noun phrases. When given the instruction, "Pick up the cube. Now put it inside the can", listeners were equally fast at identifying the intended referent for *the can* when only one of two visually available cans could accommodate the cube as when the referent was the only object in the display with that name. However, this measure cannot directly assess whether a modified noun phrase (e.g., "the big can or the can on the left") would have been more natural or effective for distinguishing the intended referent. Because referential effects on syntactic processing critically depend on expectations about modification, the current study provides a powerful test of this question. If modification were typically expected in two-referent contexts, listeners should not have been garden-pathed when only the intended referent possessed appropriate affordances for the required action. Thus, action-defined domains do not simply facilitate the identification of referents, they provide essential limits on the extent of the context used to evaluate referential noun phrases.

Our findings also show that domains are constrained to the same degree when semantically weak predicates (e.g., *put*) are used in the presence of certain tools or instruments as when semantically rich predicates (e.g., *pour*) are used. This outcome provides additional evidence that domains of interpretation are not narrowed on the basis of linguistically encoded information such as selectional restrictions. Instead, this process is driven by the listener's (nonlinguistic) conception of the action evoked by the predicate, given characteristics of the particular situation, such as the specific affordances of objects and the presence of instruments.

There are, however, many outstanding questions concerning the finer details of how object affordances and sentential constraints are coordinated to define linguistically relevant domains. For example, it is often proposed that affordances may be identified automatically in the course of object perception (e.g., Gibson, 1977). If so, the process of circumscribing domains in visual environments could be the product of two largely independent processes. For example, a restricted domain might first be defined in terms of the actions afforded by scene objects. This domain could then be narrowed by linguistic information, perhaps by computing the intersection of the affordance-based domain with the set evoked by a predicate.

We suspect, however, that domains of interpretation are the product of more continuous interactions between environment-based and communicated information sources. As a starting point, it is unlikely that all affordances for an object are calculated when it is perceived. Although a sandal can make a useful doorjamb, this property is unlikely to be noted when browsing in a shoe store. What makes particular affordances salient are specific actions or events whose identity or relevance will often be communicated by linguistic means. Thus, instead of simply directing attention within domains that have been narrowed on the basis of perceived object affordances, linguistic information may play a central role in cueing attention to relevant affordances at the outset.

Although our findings cannot be used to evaluate these alternatives, the possibility of reliably predicting actions from object affordances in our experiments was very slim given the numerous filler instructions that required many different kinds of actions (sometimes rather unusual) to be performed with scene objects. In addition, anecdotal evidence from participants' (unsolicited) comments suggests that the scene information alone was not highly predictive for the affordances that were relevant to the instruction. For example, following the trial with the display shown in Figure 1A, 1 participant remarked that she had assumed the instruction would require her to crack the (solid) egg into the empty bowl. This minimally suggests that the participant had not simply preclassified scene objects according to whether they were pourable in their given state. Nevertheless, more direct experimentation is clearly necessary to clarify how appropriate action-relevant properties are recognized during language understanding.

Modularity in Syntactic Processing

Although there are remaining questions about whether the relevant affordances are recognized before an utterance begins, the current study provides a clear test of whether this information is used at the point where a grammatical ambiguity is encountered. As noted above, previous evidence has shown that factors such as referential context and plausibility constraints affect on-line syntactic decisions. However, as we argued earlier, these effects could have arisen from domain-specific processing mechanisms that incorporate linguistic representations of discourse or form-based co-occurrences. Moreover, although previous results have established rich interactions between visual scenes and language processing, including effects of action-based affordances (cf. Chambers et al., 2002; Kamide, Altmann, & Haywood, 2003), the current results go beyond these studies in demonstrating that considerations of action can influence the earliest moments of syntactic processing—a core component of linguistic processing. Specifically, the garden-path effect typically observed when only one referent is present was reinstated when two referents were present, but only one had affordances that were compatible with the evoked action. Thus, the domain relevant for interpreting *the egg* in, "Pour the egg in the bowl over the flour," is not a linguistic–semantic representation that simply captures the number of candidates meeting the description *egg.* Rather, the domain is contoured by a nonlinguistic evaluation of how to perform an action given the affordances of referential candidates. Crucially, the preference to treat a prepositional phrase as a modifier or a goal depends on this domain. It is important to note that affordances are not likely omitted from linguistic representations of discourse context simply because of oversights in semantic theory. As argued above, the relevance of these properties cannot be established without reference to ongoing actions, which are themselves recognized on the basis of linguistic and contextual information. In summary, then, if modular systems are characterized by the rapid and automatic use of domain-specific information sources and isolation from factors such as visual perception, intentions, and the evaluation of possible actions, our results suggest that the syntactic processing system does not meet these criteria.

We can think of two lines of argument that might be available to theorists who wish to maintain an architecture that conforms to Fodor's (1983) notion of an encapsulated input system. One line of argument focuses on concerns about the visual world paradigm. The results of our and other visual world studies (e.g., Spivey et al., 2002; Tanenhaus et al., 1995; the adult subjects in Trueswell et al., 1999) showed that referential factors appear to completely eliminate the goal bias for ambiguous prepositional phrases. In contrast, studies using other paradigms such as reading have shown clear evidence that referential effects can be attenuated by lexico–grammatical information such as subcategorization preferences and co-occurrence frequencies among verbs and complement structures (see, e.g., Britt, 1994; MacDonald et al., 1994; Tanenhaus & Trueswell, 1995). This raises the question of whether the combination of a circumscribed visual world and a limited set of instructions might overestimate effects of context and mask linguistically based effects. We suspect this is unlikely for two reasons. First, a comparison of the linguistic materials in these studies reveals some important differences. Consider, for example, the study by Britt (1994), which showed that referential context eliminates the goal bias for an ambiguous phrase when a goal argument is an optional complement of the verb (e.g., "Susan dropped the book on the Civil War onto the table") but not when it is obligatory, as with the verb *put* (e.g., "Susan put the book on the. . ."). A critical feature of these sentences is that the preferred "location" sense of the preposition is possible when it introduces a goal phrase but not a modifier phrase (e.g., *on the Civil War*). Thus, both the verb *put* and the preposition *on* strongly bias the unfolding ambiguous phrase toward a goal analysis. In contrast, our experiments and the visual world studies cited above used sentences in which the location sense of *on* is plausible with both the goal and the modifier interpretation of the ambiguous phrase, thereby reducing the strong goal bias. A second point is that clear interactions of referential and lexical biases have been observed in visual world studies when these biases are explicitly manipulated (e.g., Snedeker, Thorpe, & Trueswell, 2001; see Tanenhaus & Trueswell, in press, for additional discussion). Thus, although more research clearly is needed to understand the similarities and differences in the constraints provided by written discourse and visual scenes, it seems unlikely that subtle effects of bottom-up linguistic information are masked by the use of visual contexts.

A second line of argument is that modularity might be preserved if the syntactic representations relevant to ambiguity resolution did not fully determine a particular structure. Recent work has suggested that the representations computed during on-line processing may be relatively underspecified (e.g., Ferreira, Ferraro, & Bailey, 2002). If so, structurally defined preferences may be visible only with particular classes of ambiguities or in the absence of strong nonsyntactic constraints. Our results cannot rule out this interpretation. However, garden-path effects with goal arguments are among the best documented effects in the field, and obligatory arguments of verbs are among the least likely candidates for underspecified representations. Moreover, it is difficult to understand how a mechanism of this sort would alleviate the "computational bottleneck" problem that the modularity hypothesis was originally intended to address. If syntactic information has primacy in processing, yet is weak to the extent that other information sources predominate in structural decisions, the result is largely comparable with a system with free interaction of information. In contrast, the results of our experiments, and other recent work in the field, point to a solution to the bottleneck problem that is not cast in terms of mental architecture. Specifically, the information deemed relevant for syntactic decisions appears to be limited by higher level considerations such as the recognition of communicative intentions and expectations about how action and events unfold in the world. As shown by the experiments described above, the knowledge that a referential competitor is present has different syntactic effects depending on whether the competitor can be construed as a possible participant in the action that constitutes the goal of the utterance. Thus, in contrast to the assumptions underlying the modularity thesis, the intentions and knowledge of comprehenders may form part of the solution for the bottleneck problem rather than an undifferentiated source of difficulty. It is of interest to note that this approach has recently proven useful for addressing similar kinds of questions, such as why only certain meanings are considered in the course of word learning (e.g., Baldwin, 2000; Tomasello, 2001), a process that was previously explained in terms of invariant built-in constraints (e.g., Markman, 1990). We believe it will be fruitful to consider further how the real-time coordination of information could be constrained by goals and knowledge related to the ongoing behavior of language users rather than by mechanisms that impose delays in when information becomes available.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247–264.

Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191–238.

Baldwin, D. A. (2000). Interpersonal understanding fuels knowledge acquisition. *Current Directions in Psychological Science, 9,* 40–45.

Barsalou, L. (1999). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes, 28,* 61–80.

Binder, K. S., Duffy, S. A., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language, 4,* 297–324.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure. *Journal of Memory and Language, 33,* 251–283.

Burgess, C., & Lund, K. (1997). Modeling parsing constraints with high-dimensional space. *Language and Cognitive Processes, 12,* 177–210.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47,* 30–49.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49,* 317–334.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25,* 257–271.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences, 3,* 115–120.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 320–358). Cambridge, UK: Cambridge University Press.

Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11–15.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 601–681). Hillsdale, NJ: Erlbaum.

Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge, MA: MIT Press.

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence processing. *Trends in Cognitive Sciences, 2,* 262–268.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting & knowing* (pp. 67–82). Hillsdale, NJ: Erlbaum.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43,* 379–401.

Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science, 22,* 389–424.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases.* Published doctoral dissertation, Garland, New York.

Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156.

Kamp, H., & Reyle, U. (1993). *From discourse to logic.* Dordrecht, The Netherlands: Reidel.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14,* 57–77.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes: Special Issue on Lexical Representations in Sentence Processing, 12,* 137–176.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426.

Snedeker, J., Thorpe, K., & Trueswell, J. C. (2001). On choosing the parse with the scene: The role of visual context and verb bias in ambiguity resolution. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 964–969). Edinburgh, Scotland: Erlbaum.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45,* 447–481.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12,* 153–156.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995, June 16). Integration of visual and linguistic information during spoken language comprehension. *Science, 268,* 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Speech, language, and communication* (pp. 217–262). San Diego, CA: Academic Press.

Tanenhaus, M. K., & Trueswell, J. C. (in press). Eye movements as a tool for bridging the language-as-product and language-as-action traditions. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to processing world-situated language: Bridging the product and action traditions.* Cambridge, MA: MIT Press.

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 132–158). Cambridge, UK: Cambridge University Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73,* 89–134.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33,* 285–318.

(*Appendixes follow*)

Appendix A

Supplementary Data From Experiment 1

Measure	Instruction: Ambiguous	Instruction: Unambiguous
Percentage of trials with saccade to false goal		
Incompatible competitor	47%	24%
Compatible competitor	25%	23%
Percentage of trials where saccade to false goal precedes saccade to true goal		
Incompatible competitor	31%	21%
Compatible competitor	18%	16%
Average length of individual fixation to false goal (ms)		
Incompatible competitor	556	395
Compatible competitor	319	412

Note. All values refer to eye movements occurring within the first 2,500 ms following the onset of the noun in the first prepositional phrase (e.g., *bowl* in *Pour the egg (that's) in the bowl over the flour*).

Appendix B

Supplementary Data From Experiment 2

Measure	Instruction: Ambiguous	Instruction: Unambiguous
Percentage of trials with saccade to false goal		
Incompatible competitor	31%	15%
Compatible competitor	14%	22%
Percentage of trials where saccade to false goal precedes saccade to true goal		
Incompatible competitor	29%	15%
Compatible competitor	14%	17%
Average length of individual fixation to false goal (ms)		
Incompatible competitor	656	290
Compatible competitor	367	474

Note. All values refer to eye movements occurring within the first 2,500 ms following the onset of the noun in the first prepositional phrase (e.g., *folder* in *Put the whistle (that's) on the folder in the box*).

Received May 15, 2003
Revision received September 14, 2003
Accepted October 10, 2003 ■