A Mental Spaces Analysis of the Answering Machine Paradox: 
The Agent of the Context of Utterance Revisited

Yoko Mizuta

International Christian University, Japan
mizutay@icu.ac.jp

Kaplan (1989a) claims that sentence P (I am here now) is logically true – i.e. true in every context of utterance. Its negation ~P (I am not here now) would then be logically false. Vision (1985), however, argued that ~P uttered in a voice message apparently stands for a true proposition. In the literature, this puzzle is called the Answering Machine Paradox (AMP) and has received the continuous attention of philosophers. Some scholars identify the reference of ‘I’ in the recording context. Others argue that the reference of ‘I’ is determined by elements such as the speaker’s intention, the linguistically competent and attentive audience, and the convention in which an answering machine is used. In any case, previous works hold, just like Kaplan, that the agent of ~P (and P) should be an individual and therefore that there is no agent in the playback context. In this position, the case of AMP would then be dismissed as an ‘improper utterance’, which goes beyond the scope of Kaplan’s theory.

This paper provides a novel analysis of this philosophical issue from a cognitive scientific perspective, using the mental spaces framework. AMP is explained in two steps. The key proposal made in the first step is to introduce the agent of the utterance with one’s logical identity but without his/her physical identity (Agent_log). This notion of Agent_log is given a theoretical basis in terms of a cognitive blend of mental spaces. In the second step, the

---

1 An earlier version of this paper was presented at the International Conference on Language, Communication and Cognition (Brighton, August 4-8, 2008). The author is grateful to the audience for helpful comments. Thanks also go to anonymous reviewers for insightful and helpful comments and to Katie Gruber for proofreading the paper. Any remaining shortcomings are, of course, mine.
voice message involves a pragmatic mapping from the Agent_log to a relevant individual, who is the ultimate reference of ‘I’.

The apparent puzzle in question is attributed to Kaplan’s empirical assumption that the agent should be an individual, as in a face-to-face communication. With the introduction of the notion of Agent_log, a wider range of messages can be accommodated as proper utterances and Kaplan’s theory of indexicals is essentially defended.

**Keywords:** Indexicals, the answering machine paradox, agent of the context of utterance, mental space theory, pragmatic reference

1. Introduction

Kaplan (1989a) proposed the semantics of pure indexicals (‘I’, ‘here’, and ‘now’) in such a way that ‘I’, ‘here’, and ‘now’ refer respectively to the agent, the location, and the time of the context of utterance. (The function which assigns indexicals their semantic values with respect to a context is called ‘character’. He distinguishes ‘character’ from ‘content’, the actual reference of indexicals). He limits himself to the semantics of indexicals in ‘proper utterances’, in which the agent is located at the place and time of utterance. Kaplan claims that sentence (1) is logically true – i.e. true in every context of utterance. Its negation (2) would then be logically false.

1. I am here now.
2. I am not here now.

Vision (1985) as well as Kaplan himself (Kaplan 1989a, p.491, fn12), however, argued that (2) uttered in a voice message apparently stands for a true proposition. In the literature, this apparent puzzle is called the Answering Machine Paradox (AMP) and has received the continuous attention of philosophers, concerning its effect on Kaplan’s theory of

---

2 ‘Let the class of indices be narrowed to include only the proper ones – namely, those \([w, x, p, t]\) such that in the world \(w, x\) is located at \(p\) at the time \(t\). Such a move may have been intended originally since improper indices are like impossible worlds; no such contexts could exist with respect to them’ (Kaplan 1989, p.509).
Previous works, just like Kaplan himself, hold that the agent of a message should be an individual and therefore that there is no agent in the playback context, which is presumably the context of interpretation and where the references of ‘here’ and ‘now’ are determined. In this position, the case of an answering machine would then be dismissed as an ‘improper utterance’, which is simply beyond the scope of Kaplan’s thesis.

This paper provides a novel analysis of this philosophical issue from a cognitive scientific perspective, using the mental spaces framework. The aim of the paper is two-fold: On the one hand, it provides an account of AMP, and, on the other hand, it explains its implication for Kaplan’s theory.

The structure of the paper is as follows. I will start by making some preliminary discussions (Section 2), and clarifying my approach in relation to previous work (Section 3). I will then introduce the mental space theory (Section 4) and provide a mental spaces analysis of AMP (Section 5). With discussions of related issues (Section 6), I will conclude the paper by defending Kaplan’s theory—specifically, the ‘character’ of indexicals and the thesis that (1) is logically true—, with an additional thesis about the agent of the context of utterance (Section 7).

2. Preliminaries

2.1 The Deferred Utterance Intuition

Let us consider a concrete example. Suppose that John left the following message on his answering machine in his office before traveling.

(3) Hello. This is John. I’m sorry, I am not here now. Please leave a message.

In the recording, John speaks as though he is talking to a future caller. (Technically, of course, John does not need to have any caller in mind. However, if he imagines a caller, the message will sound more natural.) Importantly, the underlined part (i.e. (2)) does not describe the situation in the recording context. Instead, it is intended to describe the situation of a
later call.

In the playback, the message comes into effect in response to each phone call. The underlined part conveys that John is not in his office at the moment of the call (because he is traveling). In Sidelle’s (1991) term, it becomes a ‘genuine utterance’. Sidelle analyzes a recorded message case as ‘arranging to make an utterance at a later time, or, if one likes, deferring an utterance’ (p.535). This intuition is called the ‘deferred utterance intuition’ (DUI).

Stevens (2008, p.217) argues against DUI. He claims that there is an utterance in the recording context and there is no utterance in the playback context. He puts it: ‘one can only record an instance of (2)\(^3\) on an answerphone by *uttering* (2)’ (with original italics).

It is true that one speaks in the recording. (Sidelle actually does *not* say that one does not speak in the recording.) Sidelle’s point is that the utterance made in the recording context *comes into effect* later in the playback context. If one wishes, s/he may interpret that there are two ‘utterances’; the original one in the recording context and the deferred one in the playback context. A related point to mention is that Stevens seems to limit the notion of an utterance to an individual’s production of sounds, as in ordinary utterances and in the recording context. In contrast, Sidelle apparently extends the notion of an utterance to include an individual’s pre-recorded message played back by a machine.

Furthermore, if no utterance is acknowledged in the playback context, no distinction would be made between an authorized playback in response to a call and a test playback, which is in fact a mere production of the sound with no communicative purpose or effect.

Cohens (2013) supports DUI:

Sidelle’s ingenious proposal is that answering machines, notes, and the like are utterance-deferral technologies – technologies that permit one to produce in a context \(c^i\) an indexical-containing string that will be interpreted *not* with respect to \(c^i\) (as usual), but with respect to some distinct context \(c^t\), in which that string will

\(^3\) ‘I am not here now’.
eventually be tokened. (p.7)

In light of DUI, I extend the notion of an utterance to include various occurrences of linguistic expressions for communicative purposes. In the case of an answering machine, a plausible view would be that there is an utterance in both the recording and the playback contexts. They have different properties. The one in the recording context is an individual’s production of a message with no communicative effect at that moment. The one in the playback context is a machine’s production of an individual’s pre-recorded message with a real-time communicative force triggered by each phone call.

From a communicative perspective, the focus is on the playback context (i.e. the ‘context of interpretation’), where the indexicals are assigned their semantic values. Accordingly, the reference of ‘here’ and ‘now’ are assigned their values respectively to the location and time of the playback context.

Now, what about ‘I’? — What is the reference of ‘I’ and how is it determined? These questions play a central role in the explanation of AMP and I will come back to these questions later.

2.2 The Agent of the Context of Utterance vs. An Utterer\(^5\)

I would also like to comment on the notions, the ‘agent of the context of utterance’ vs. an ‘utterer’. In Kaplan’s (1989a) theory, the character of ‘I’ is the ‘agent of the context of utterance’.

An utterer is concerned with the production of the sound. Thus, in the case of a voice message, the one who records the message is the utterer. One may consider that an answering machine is also possibly an utterer in the playback context. However, the agent of the context of utterance seems to necessitate a different role.

---

\(^4\) Cohen (2013) claims that ‘I’ does not shift its semantic referent in the way that ‘here’ and ‘now’ apparently do in answering-machine uses and that the referent of ‘I’ is the individual in the recording context (p.3). But this is controversial, as we will see in section 3.1.2.

\(^5\) I will use the term ‘utterer’ in the sense of ‘one who produces a/the utterance’, although a native speaker said the term (vs. ‘speaker’) sounds awkward to her.
Goffman (1981, pp.144f.) discusses the different roles of a speaker, which are ‘Animator’ (‘individual active in the role of utterance production’), ‘Author’ (‘someone who has selected the sentiments that are expressed and the words in which they are encoded’), and ‘Principal’ (‘someone whose position is established by the words that are spoken, someone whose beliefs have been told, someone who is committed to what the words say’).

Kaplan assumes ordinary (i.e. real time, face-to-face) utterances where the same person inhabits all these speaker roles. But when we analyze the case of an answering machine, we need to distinguish between the different roles. I consider that the utterer corresponds with the Animator role, whereas the agent of the context of utterance corresponds with the Principal role. Thus, for example, if Bob records a message to be used on John’s answering machine, the utterer is Bob, whereas the agent of the context of utterance involves John.

3. Previous Work and My Approach

In this section, I will overview previous work concerning two specific points, and then outline my approach.

3.1 Should the Agent of the Context of Utterance be an Individual?

3.1.1 An assumption in the literature

Kaplan (1989a) assumes that the agent of the context of utterance is an individual, in view of ordinary, face-to-face communication. The literature shares this assumption, even when considering the case of an answering machine and related cases. What is the consequence of this assumption? It is that the case of an answering machine would be reduced to an ‘improper utterance’, given that there is no individual in the playback context, which is the plausible context of utterance (as discussed in section 2.1). Remember that Kaplan is only concerned with ‘proper utterances’, and his thesis that

---

6 McCawley (1999) discusses the relation between ‘I’ and participant roles.
7 Romdenh-Romluc (2006, p.259) mentions a similar example.
(1) is logically true concerns proper utterances (p.509). With the above-mentioned assumption, we see that AMP simply goes beyond the scope of Kaplan’s theory. That is, the observation that the proposition (2) is true in the case of an answering machine would not affect Kaplan’s thesis.

Predelli (2011) presumably has the same view. In his conclusion he writes: ‘a variety of views regarding written notes, recorded messages, and the characters of ‘I’, ‘here’, and ‘now’ yield a satisfactory explanation of the problem under discussion to the extent to which they accept improper contexts…….the evidence provided by written notes and recorded messages should not constrain the shape of empirically satisfactory semantic account of indexicals’ (p.302, with my italics).

3.1.2 An overview of previous work

The goal of previous work seems to have been to explain the case of an answering machine as such and/or to explore a theory of indexicals which covers non-ordinary (‘improper’) utterances like the cases of an answering machine and written notes. I will briefly introduce the three major lines of analyses in the literature. (For detailed discussions of major semantic and pragmatic approaches, see Cohen & Michaelson 2013.)

The first line of analysis (Sidelle 1991; Whitsey 2003; Cohen 2013; Cohen & Michaelson 2013: ‘the context of tokening view’) is semantic and proposes a minimum modification of Kaplan’s original theory. It supports DUI and claims that the reference of ‘here’ and ‘now’ shift to the playback context. On the other hand, it claims that the agent does not shift and remains in the recording context. This view then attributes the truth of (2) to the discrepancy between the two contexts — the recording context, where (in their view) ‘I’ is assigned its value, and the playback context, where ‘here’ and ‘now’ are assigned their values. This analysis of (2), however, seems to be counterintuitive, given that the proposition (2) concerns the location of the speaker (say, John) at the time of playback, not at the time of recording. The proposition (2) is true iff the speaker John is not in the playback context (e.g. while he is traveling).

The second line of analysis (Michaelson 2014: ‘the character-shifting view’) is also semantic. It acknowledges some ‘context types’ including the
face-to-face context and argues that the character of indexicals shifts from
the original, Kaplan’s version (for face-to-face contexts) in the case of other
context types. He then posits the character for each context type (i.e. face-
to-face, answering machine, and postcard). For the answering machine
case, the character is posited as follows (p.528, part of his (8)):

Answering Machine

‘I’ refers to the owner of the line
‘here’ refers to the location of playback
‘now’ refers to the time of playback

On this view, ‘The conventions that arise out the use of recording
technologies can shift the character of indexicals from their original
character’ (p.538, with original italics) and recorded language is regarded as
‘quite literally an extension of spoken language’ (p.539).

The third line of analysis is given mainly from a pragmatic perspective.
that ‘written and recorded messages are to be evaluated with respect to the
intended context of interpretation’ (Predelli 1998) and that sometimes the
context of interpretation does not coincide with the context of utterance. Others argue that the reference of ‘I’ is determined by elements such as the
conventional setting in which a recorded message is used (Corazza, Fish &
Gorvett 2002: ‘the conventionalist view’), or the linguistically competent
and attentive audience (Romdenh-Romluc 2006: ‘the recognized context
view’). Note, however, that Cohen (2013) mentions that Kaplan’s theory
is about the semantics of indexicals, rather than their pragmatics, and
that ‘we are independently committed to distinguishing speaker reference
from semantic reference, and to thinking that conveyed content can fail to
coincide with semantic content’ (p.7).

Now, I will stop and reconsider the alleged assumption that the agent is
an individual.

8 Predelli’s analysis can be regarded as semantic, for it mentions the context of evaluation
of indexicals. However, I would think that it can also be regarded as pragmatic, for we
could consider, following Cohen and Michaeson (2013, p.585), that it apparently concerns
speaker reference, as opposed to semantic reference. I will leave this semantic/pragmatic
difference open in this paper.
3.1.3 Reconsidering the alleged assumption

Notice that in the voice message in the playback, various illocutionary forces (Austin 1975) are acknowledged. For example, ‘I’m sorry’ has the illocutionary force of an apology, whereas ‘This is John’ and ‘I am not here now’ have the illocutionary force of an assertion. And ‘Please leave a message’ has the illocutionary force of making a request. Specifically, ‘I’m sorry’ is essentially different from ‘John is sorry’: In the former, the speaker apologizes while uttering it, whereas in the latter the speaker describes a certain state. Even though it is a machine which produces the sound, illocutionary forces are recognized in the playback at each call, as far as the voice message is taken seriously in real time.\(^9\)

In performative utterances (Austin 1975), the speaker does something (e.g. makes an apology, an assertion, etc.) while producing the utterance. Thus it is essential for the agent to be in the context of utterance.

Now, we are in a dilemma between two incompatible assumptions:

(4) Incompatible assumptions about the agent of the context of utterance

1. The agent of the context of utterance is an individual.  
   (This is mentioned in the previous section. It is Kaplan’s assumption in view of ordinary, face-to-face communication and the assumption shared in the literature.)

2. The agent of the context of utterance should be in the context of utterance (in our case, in the context of playback), from a speech act perspective.

In the case of a voice message, the above assumptions are incompatible with each other, since there is no individual who answers the call. The question then is which assumption is more central.

The first assumption above is only empirical, rather than logical, coming from the situation in which ordinary utterances are made. Since we are now concerned with a voice message, there is no reason to make this first

\(^9\) Bianchi (2014, p.486) mentions a similar point.
assumption. Importantly, the agent of the context of utterance is the speaker with respect to the relevant ‘occurrence’ of ‘I’. Since we are assuming that the utterance is made effective in the playback, there is a good reason for considering that ‘I’ occurs in the playback, just as ‘here’ and ‘now’ do.

The first assumption is also motivated by the situation in Kaplan’s time. While face-to-face communication is apparently the default even in our time, we are invited to consider a wider variety of cases triggered by the appearance of new technologies, including an answering machine. Michaelson (2014) puts it, ‘the best way to accommodate these data is to conceive of recording technologies as introducing special, non-basic sorts of contexts, accompanied by non-basic conventions governing the use of indexicals in those contexts’ (p.1).

The second assumption above, which is motivated by Austin’s (1975) speech act theory, is more central. It also plays an essential role in distinguishing an actual voice message from a text playback. I thus suggest that we adopt the second assumption and discard the first.

3.1.4 An option proposed by Voltolini (2006)

Voltolini (2006) says that in appropriate fictional settings the agent of the context of utterance may be a fictional object. He explains relevant cases in terms of a make-believe game as a socially shared activity.

Regarding the time parameter of the case of an answering machine, he analyzes as follows: ‘The speaker of (2)[I am not here now] is playing a make-believe game in which he is pretending to be not at the time in which he is speaking – 2 p.m. – but rather at another time, the time in which he expects his wife to hear the message, i.e., 5 p.m. Thus, 5 p.m. is the pretended time of that utterance of (2)’ (p.28, with my note). This analysis goes along the lines of Sidelle (1991) (see p.3).

Regarding the agent parameter, Voltolini analyzes the following example of a voice message:

(5) I am not myself, but my answering machine. (Voltolini 2006, p.35, his (14))
He argues: ‘that occurrence of “I” does not refer to that original utterer, but rather to the answering machine itself!’ (p.36, with original italics) and ‘as far as the occurrence of “I” is concerned, the original utterer of (14) pretends to be something else; namely, his own answering machine. So, my answering machine is the pretended agent of the utterance of that token of “I” ’ (p.36). He illustrates that I refers to the answering machine in a certain fictional setting.10

Voltolini’s intuition that the interpretation of a voice message may involve a fictional context will be of some relevance to my analysis.

3.1.5 Summary

I suggest that we take the second assumption in (4), discarding the first. This means that the case of an answering machine and the like are included in proper utterances. Therefore, AMP is now within the scope of Kaplan’s theory. We thus need to explain it and clarify its implication for Kaplan’s theory. The key question ‘How can the agent exist in the context of playback?’ will be considered in section 5.1.

3.2 On the Phenomenon of ‘Reference-Shifting’

The other issue I’d like to consider in this section is the phenomenon of ‘reference-shifting’. Mount (2008) observes that pure indexicals do not automatically refer to the agent, the location, and the time of the context of utterance, even in ordinary utterances. That is, he challenges the character of indexicals proposed by Kaplan. Let us focus on ‘I’.

(6) [Uttered in a game of Monopoly] I am on a purple square. (Mount 2008, p.200)
(7) [A car owner says] I am (parked) out back. (Mount 2008, p.200)

10 Regarding I in (2), however, he argues that it refers to the original utterer, as he says: ‘In (14), the occurrence of “myself” (as well as the occurrence of the possessive “my”) is not problematic: it refers to the original utterer of the sentence, as well as “I” does in (2)[I am not here now]’ (p.36, with my italics and note). Sentence (2) produced by an answering machine will then be a case of an improper utterance.
In Mount’s view, ‘I’ in (6) and (7) do not refer to the speaker: They refer respectively to the shoe-shaped game piece on the board and the speaker’s car. What implication do these data have for the semantics of ‘I’?

Åkerman (2015) discusses examples similar to (8) below. He observes that ‘The referent of ‘I’ appears to shift from Mrs Horne in the first occurrence to her game piece in the second’ (p.7).

(8) [Uttered by Mrs. Horne during a game of Monopoly] I know that I am on a Baltic Avenue. (p.7, with my boldfaces)

For such data, Åkerman proposes a pragmatic account in the framework of Recanati (2004). He starts with the following example given by Nunberg.

(9) The ham sandwich has left without paying. (Åkerman, 2015, p.24)

This is a typical example of metonymy: ‘The ham sandwich’ refers to the person who ordered the ham sandwich. Åkerman’s quotes Recanati’s (2004) account of (9):

[T]he description ‘the ham sandwich’ first receives its literal interpretation in such a way that a representation of a ham sandwich is activated; activation then spreads to related representations, including a representation of the man who ordered a ham sandwich. All these representations activated by the description ‘the ham sandwich’ contribute potential candidate for the status of the semantic value of the expression; …… Now, the ham sandwich orderer is a better candidate than the ham sandwich itself for the status of argument for ‘has left without paying’. It is therefore the derived, non-literal candidate which

---

11 For a different view proposed by Nunberg (1993) and relevant discussions by Mount, see Mount (2008, pp.200-202).

12 For a critical overview of arguments against pragmatic approaches, see Åkerman (2015, pp.9-22).
is retained, while the literal interpretation is discarded (Recanati, 2004, p.29; Åkerman, pp.24f., with original italics).

The shift from the literal interpretation to a non-literal interpretation is called ‘transfer’ in Recanati’s (2004) theory. Åkerman applies this pragmatic account\(^\text{13}\) to cases like (8):

> All occurrences of the pronoun ‘I’ first receive their literal interpretation, in such a way that a representation of the speaker is activated; activation then spreads to related representations, including a representation of the speaker’s game piece. (p.28)

He also mentions that ‘any candidate referent for ‘I’ must be appropriately related to the speaker in order to be genuinely available as a pragmatic referent’ (p.30). In more general terms: ‘the process leading to a transfer turns on there being an appropriate relation between the literal and the non-literal referent’ (p.30).

Åkerman’s account applies to Mount’s (2008) examples (6) and (7) too. We could consider that in (6) and (7), ‘I’ first refers to the speaker, and then, by way of transfer, refers respectively to the speaker’s game piece and to the speaker’s car. Then, the reference-shifting in (6) and (7) does not affect the semantics of ‘I’.

In summary, Åkerman illustrates cases in which referent-shifting of ‘I’ can be explained pragmatically, without modifying the character of ‘I’ proposed by Kaplan.

### 3.3 My Approach

Based on what has been discussed so far, I will take the following approach in my analysis:

\(^\text{13}\) Although Recanati’s (2004) account is concerned with the status of the semantic value of the expression, his account is still regarded as pragmatic, for it crucially requires a pragmatic link between the literal and the non-literal references.
1. In light of DUI, I consider that in the case of an answering machine, the playback context is the context of utterance/interpretation.

2. I extend the notion of the ‘agent of the context of utterance’ to a non-individual so that it can exist in the playback context.

3. Accordingly, the answering machine message in the playback will be treated as a proper utterance, which is within the scope of Kaplan’s theory.

4. I will apply the character of indexicals proposed by Kaplan to the case of an answering machine and defend Kaplan’s theory — specifically, the character of indexicals and the thesis that (1) [I am here now] is logically true.

5. The puzzle at issue (i.e. the fact that (2) [I am not here now] is apparently true in the voice message) will be explained pragmatically, without a contradiction with the logical truth of (1).

To the philosophical issue of AMP, I will provide a novel analysis from a cognitive scientific perspective according to the mental space theory (in section 5).

4. Mental Space Theory

As a basis for my analysis, I will introduce the essence of the mental space theory.

4.1 Conceptual Blend

Mental space theory (Fauconnier 1994, 1997, Turner & Fauconnier 1995) is a useful device for the analysis of human thought and linguistic activities and has a wide range of application (e.g. Hiraga 1999, Veale 1999). The key notion of ‘blending’ of mental spaces applies to conceptual integration. Complex cognitive process involves multiple applications of blending.

To take an example, Fauconnier analyzes counterfactuals like the following:
The question here is why the final word should be *me*, not a reflexive *myself*, despite that it is c-commanded by the subject *I*.

Fauconnier explains the phenomenon in terms of blending of two mental spaces, as illustrated in Figure 1 (Fauconnier 1997, p.162, his Figure 6.10, with my minor modifications). One space (Input 1) is the real-world space, where the speaker (b₁ as a worker) talks to the addressee (a₁ as an employer), and the other (Input 2) is a hypothetical, counterfactual space where the speaker is in the addressee’s position (a₂ as an employer) and the addressee is a worker (b₂). The elements a₁ and a₂ as well as b₁ and b₂ are linked by an analogical function (*A*) on the basis of the position, whereas a₂ and b₁ are linked by an Identity function (*Id*) on the basis of the identity of the person.

In the blended space, there are two elements (a’ and b’), each of which partially inherits the speaker’s dispositions. Specifically, a’ inherits the speaker’s essential properties from Input 2, whereas b’ inherits his position and strengths (which belong to the addressee in the real world) from Input 1. The sentence says that a’ would hire b’. Given the discrepancy between

![Figure 1. Counterfactual blend in the sentence *If I were you, I would hire {me/*myself}*.](image-url)
the two references (a’ and b’), it is no surprise that personal pronoun \textit{me}, not reflexive \textit{myself}, is the right form. Example (10) would be a puzzle if one tries to analyze it based on a single space.

Mental space theory works well for the analysis of cases like this, which involve hypothetical situations of various kinds. The case of an answering machine involves a certain fictional setting (Cf. Voltolini 2006), and therefore the mental space theory is expected to be a useful device for its analysis.

4.2 Pragmatic Reference

Another key element of my analysis of AMP is pragmatic reference. As the literature discusses (Nunberg 1978, Fauconnier 1994, Rubba 1996), indexicals can have a variety of references on the basis of pragmatic arrangements.

Fauconnier (1994) introduces the following principle, which plays a key role in explaining a variety of pragmatically determined references.


If two objects (in the most general sense), \(a\) and \(b\), are linked by a pragmatic function \(F (b = F(a))\), a description of \(a\), \(d_a\), may be used to identify its counterpart \(b\).

Fauconnier calls \(a\), \(b\), and \(F\) in ID Principle ‘trigger’, ‘target’, and ‘connector’, respectively.

To take an example to which ID Principle applies:

(12) Plato is on the top shelf. (Fauconnier 1994, p.4)

There is a pragmatic mapping (‘connector’) between the author Plato and the book of his work. By virtue of this connector, the expression ‘Plato’ in (12) as a trigger can refer to the book as a target.

To take another example:

(13) In Len’s painting, \textit{the girl with blue eyes} has green eyes.
What sentence (13) says is not a contradiction. The expression ‘In Len’s painting’ introduces a connector between the discourse space (i.e. the real world space) and Len’s painting. The italicized expression serves as a trigger, and her counterpart (who has green eyes) becomes the target. That is, (13) means ‘The counterpart of the girl with blue eyes has green eyes’.

Rubba (1996) gives a similar example, which contains ‘I’.

(14) In this picture, I’m wearing a gorilla suit. Rubba (1996, p.233)

He explains, ‘The expression in this picture builds a mental space’. Then a connector (in Fauconnier’s sense) is pragmatically constructed between the discourse space and the picture space. ‘The connector and other semantic features inherent in the pronoun I allow a listener to use the pronoun as a trigger to identify the person wearing the gorilla suit in the picture (the target) with the speaker of the sentence in the actual discourse context’. (Ibid. p.233)

In short, ‘I’ in (14) semantically refers to the speaker, and then pragmatically refers to his counterpart in the picture. To be noted, ‘I’ does not directly refer to the one in the picture: Therefore, with (14) there is no need to modify Kaplan’s semantics (‘character’) of ‘I’.

As we see, Recanati’s (2004) theory of transfer (see section 3.2) and Fauconnier’s ID principle share the same essential idea. In fact, we can explain examples like (6) and (7), repeated here as (15) and (16), in terms of ID principle.

(15) [Uttered in a game of Monopoly] I am on a purple square. (Mount 2008, p.200)

(16) [A car owner says] I am (parked) out back. (Mount 2008, p.200)

In (15), there is a connector between the speaker and her game piece, and the expression ‘I’ for the speaker serves as a trigger and refers to the target, her game piece. In (16), there is a connector between the speaker and her
car, and the expression ‘I’ serves as a trigger and refers to the speaker’s car as the target. The difference between the two theories is the following. In Recanati’s (2004) theory, transfer provides potential candidates for the semantic reference of the expression at issue, whereas in Fauconnier’s theory, a connector (i.e. pragmatic function/ mapping) provides a pragmatic, not semantic, reference of the expression at issue.

Suppose next that John is watching a video in a living room of his house. In the video, he is also in the living room and walking around. John says:

(17) In the video, I’m not sitting here.

In (17), ‘I’ refers to John’s counterpart in the video, by way of the speaker John. On the other hand, ‘here’ first refers to a certain location in the living room, where John is. And then, by virtue of the connector, it refers to the corresponding location in the video.

The same mechanism involving pragmatic reference is at work in the case of the answering machine, as will be discussed in section 5.2.

5. An Analysis of the Answering Machine Paradox

The interpretation of the answering machine message in the playback involves two steps. Both steps are explained in mental spaces terms. In the following, I will explain each step and provide the whole picture of my analysis. For practical purposes, let us suppose that John records his message (3) for his own use.

5.1 The Structure of the Playback Context- Step 1

5.1.1 The logical identity of the agent

In an ordinary utterance, the agent is an individual who has both logical and physical identity. In a non-ordinary utterance like an answering machine message, I propose that there is an agent in the context of utterance (e.g. the playback context) only with an individual’s logical identity, lacking his/her physical identity. This is the essential difference between ordinary and non-
A Mental Spaces Analysis of the Answering Machine Paradox

ordinary utterances. Let us call such an agent Agent_log, in the sense of the ‘agent with someone’s logical identity’.

(18) Definition of Agent_log
Agent_log is an agent of the context of utterance with the logical identity of some individual but without the physical identity of that individual.

If one wishes, s/he can call Agent_log a ‘virtual agent’, in the sense that s/he lacks a physical identity.

As far as the message is taken seriously by the caller, that is, as far as communication is in effect, the playback context does not lack an agent but has an Agent_log. In the case of a test playback, in contrast, there is no agent, given that it is a mere production of the sound without communicative effects. In the case of an ordinary utterance, Agent_log is accompanied by the same person’s physical identity: The two kinds of identity together make an individual. To put it differently, we can see an individual, decomposing him/her into his/her logical and physical aspects.

In the apology ‘I’m sorry’ in (3), there is the agent of the context of utterance, who is Agent_log, and who is the reference of ‘I’.

In order to provide a better understanding of the notion of Agent_log, I will illustrate the conceptual structure of Agent_log using the mental spaces framework, in section 5.1.2 below.

5.1.2 Context blend

The first step involved in the case of an answering machine is a conceptual blend of two contexts (‘context blend’)\(^\text{14}\), which is illustrated in Figure 2. The playback context is expressed as a blended space \(S_3\). This step of context blend gives a theoretical basis for the notion of Agent_log.

\(S_1\) is a reality space in which the machine \((a_1)\) answers the phone with the voice message \((b_1)\) real time. The machine produces the sound, and the voice message has a certain content and style.

\(^\text{14}\) See Fauconnier (1997) for details of conceptual blends.
S₂ is a generic space in which a likely individual a₂ (in our example, John) answers the phone as necessary. Those who call John presumably have this generic space in mind, expecting him to answer the phone. Elements a₁ and a₂ are connected by an analogical function (A), on the basis of the phone number. John in S₂ (i.e. a₂) has both a logical and a physical identity, and what he utters (b₂) is authorized and intended by him.

The blended space S₃ is the playback context, which is created by a partial projection of S₁ and a partial projection of S₂. The arrows in the figure indicate these projections. The former projection concerns the time (t₁) /place (p₁) and the content/style of the message. The latter projection concerns the dispositions and intentions of John.

By virtue of these projections, in S₃ it sounds like John is answering the phone and authorizes the message in real time. ‘I’ in the voice message in the playback context refers to the speaker a₃ in S₃, who inherits essential properties of a₂ (John) and makes real-time utterances of the recorded message. That is, a₃ is Agent_log of the voice message. (It may be worth pointing out that the physical identity of a₃ is inherited from a₁.) The message b₃ in S₃ inherits the content and style (and vocal quality) from b₁, whereas it inherits John’s authorization and intention from b₂.

![Figure 2. Context blend in the voice message](image)

The point is that the playback context is not a single space in reality in which an answering machine produces a sound. The context involves conceptual blending of two spaces and an agent of the context of utterance
is in the blended space, as Agent_log. In the presence of this agent (Agent_log), the utterances made in $S_3$ can have illocutionary forces such as an apology (e.g. *I'm sorry*), an assertion (e.g. *This is John. I am not here now*), and making a request (e.g. *Please leave a message*). On this basis, the voice message in the playback context is now accommodated as a proper utterance. As far as ‘I’ refers to Agent_log, (1) is true in $S_3$. Then, how can (2), the negation of (1), be true? This is explained in step 2 discussed below.

### 5.2 The Structure of the Playback Context — Step 2

The second step is explained in terms of pragmatic reference introduced in section 4.2. There is a pragmatic mapping between the playback context and the real world. Agent_log in the playback context, which is the reference of ‘I’ in the first step, pragmatically refers to his counterpart in the real world, who is an individual (John). On the other hand, the references of ‘here’ and ‘now’, which are assigned their values in $S_3$ will not involve further steps. As a result, there is a discrepancy between the space where ‘I’ has its ultimate reference ($S_4$ in Figure 3 below) and the one where ‘here’ and ‘now’ are assigned their values.

### 5.3 The Whole Structure

The whole structure of AMP is illustrated in Figure 3. The first step is a conceptual blend of two spaces—the real world space in which the machine ($a_1$) answers the phone and the generic space in which John ($a_2$) answers the phone as necessary. In the blended space, there is an agent of utterance ($a_3$), who is Agent_log. The Agent_log $a_3$ is connected to $a_1$ by an analogical function ($A$) inheriting the production of sound. It is also connected to $a_2$ by an identity function ($Id$), inheriting the logical identity of John.

The second step is a mapping from $S_3$ to the real world space $S_4$, in which John is, for example, traveling at the time of playback. Sentence (2) means the same thing as (19) below: ‘in reality’ introduces a real world space and constructs a connector to it.

(19) In reality, I am not here now.
Even without the expression ‘in reality’, we can assume that there is a pragmatic mapping (‘connector’) between the playback space $S_3$ and the real world space $S_4$. This can be considered as part of the convention in which an answering machine is used and what a linguistically competent and attentive listener can recognize (Romdenh-Romluc 2006). By virtue of this connector, speaker $a_3$ (Agent_log) in $S_3$ becomes a trigger and his counterpart in $S_3$ becomes a target. In other words, $a_3$ is pragmatically connected to John in $S_4$ by an identity function ($Id$). As a result, ‘I’ can refer to John as an individual.

In (2), ‘I’ ultimately refers to John in $S_4$, whereas ‘here’/’now’ refer to the place/time of $S_3$.

![Figure 3. The whole structure of AMP](image)

To summarize, ‘I’ in (2) first refers to Agent_log ($a_3$) within the playback context ($S_3$) on a semantic basis, and then, it pragmatically refers to the individual John ($a_4$) in the real world $S_4$. On the other hand, the references of ‘here’ and ‘now’ are assigned their values within $S_3$. Thus, the truth of (2) is attributed to the discrepancy between $S_3$ and $S_4$ (i.e. between $a_3$ and $a_4$).
Since ‘I’ semantically refers to the agent of the context of utterance, Agent_log, the semantic references of ‘I’, ‘here’, and ‘now’ are all assigned their values in the playback context. So far as ‘I’ refers to Agent_log, (1) is true and (2) is false. However, once pragmatic reference is involved in the second step, (2) becomes true: It illustrates that pragmatics contributes to the truth condition of a sentence (so-called “pragmatic intrusion”). Since the truth of (2) is attributed to pragmatics, there is no need to modify the semantics (i.e. the character) of ‘I’ in Kaplan’s theory: ‘I’ refers to the agent of the context of utterance. Kaplan’s theory — specifically, the character of indexicals and the thesis that (1) is logically true — is thus defended.

6. Related Issues

Now let us consider some examples in relation to the proposed analysis of AMP, especially to the notion of Agent_log.

6.1 Ordinary Utterances

In the previous section, the case of an answering machine was analyzed in such a way that ‘I’ semantically refers to Agent_log. Then, what about ordinary utterances (i.e. those in which an individual utters in real time, face-to-face)?

There are two ways to explain this. The first approach follows Kaplan’s assumption: ‘I’ refers to the individual, the speaker, as the agent of the context of utterance. The second approach involves Agent_log as in the case of a voice message: 1) ‘I’ refers to the speaker’s logical identity (Agent_log), and 2) via a trivial connector from the speaker’s logical identity to his/her physical identity, Agent_log is then mapped onto the whole individual.

The first approach sounds simple, but according to my analysis it is actually a conventionalized version of the second approach. Regarding the second approach, we could consider that since the logical and physical aspects of the speaker are combined in this case, we don’t decompose the process into two steps but capture the whole process as a single step. The first approach comes from the conventionalization of this ‘single’ step. In other words, the intuition that the agent of the context of utterance is an
individual involves pragmatics.

In this light, I argue that the reference of ‘I’ is semantically the logical identity of the individual at issue (i.e. Agent_log), even in the case of ordinary utterances. Although we are familiar with ordinary utterances, we could consider that an ordinary utterance is rather a special case of a wider variety of utterances in that the individual’s physical identity accompanies his/her logical identity.

In order to incorporate this idea into the theory of indexicals, I propose to add the following thesis:

(20) A thesis about the character of ‘I’:

‘The agent of the context of utterance’, which appears in the character of ‘I’ in Kaplan’s theory, means a logical agent (Agent_log).

This thesis explicitly denies Kaplan’s (and the literature’s) assumption that the agent of the context of utterance is an individual. Putting the character of ‘I’ (i.e. ‘I’ refers to the agent of the context of utterance) and the thesis (20) together, the semantic reference of ‘I’ proves to be consistently Agent_log, in both ordinary and non-ordinary utterances. This is the modification I make to Kaplan’s theory.\(^{15}\)

6.2 Comparison with a Fiction

The case of a voice message has both a similarity to and a difference from fictional cases.

Suppose that Macbeth speaks in a play. The whole play is in a fictional setting. In Macbeth’s utterance, the agent is Macbeth as a fictional individual, without a connection to the real world space in which the audience is watching the play.

In the case of a voice message, the agent (Agent_log) has only a logical identity, as I argued. This point is similar to a fictional case like a play.

\(^{15}\) One may prefer to directly modify the character of ‘I’ to: ‘the reference of ‘I’ is Agent_log’. But I chose to maintain Kaplan’s character itself and posit this additional thesis.
There is a crucial difference between the case of a voice message and a fictional case, though. In the former case, the message is real, not fictional. The listener is expected to take it seriously. In the play, in contrast, the whole space is fictional and independent of the audience’s space.

### 6.3 A Voice Message Recorded by Somebody Other Than the Agent

Romdenh-Romluc (2006, p.259) discusses a case that Penelope records a message (2) on behalf of Kaori to be used in Kaori’s office. She correctly claims that the agent of utterance is Kaori, not the utterer Penelope, on the basis that a linguistically competent and attentive listener would understand the message this way.

My analysis proposed in section 5 works perfectly in this case too. A crucial element is the Input space 2 illustrated in Figure 2 (a generic space in which John — in this case, Kaori — answers the phone as necessary). Those who call Kaori would expect that she answers the phone, and therefore the agent will be identified with Kaori.

In my model, regardless of who recorded the message — even in the case of computer-synthesized messages — the agent will be regarded as the logical identity of Kaori because the playback space is partially projected from the generic space where Kaori answers the phone. For the correct identification of the agent, it is crucial that the listener involves this generic space in her interpretation. This point can be attributed either to the linguistic competence and attentiveness of the listener (Romdenh-Romluc 2006), or to the convention in the use of an answering machine (Corazza, et al. 2002, Michaelson 2014).

### 6.4 Another Example of a Voice Message

The following is another example of a voice message on an answering machine.

(21) Hello. I cannot come to the phone right now. Please leave a message
It is an apparent puzzle that the caller hears John’s voice on the phone, who says he ‘cannot come to the phone right now’. We could analyze that Agent_log with John’s logical identity (and the physical identity of an answering machine) answers the phone. Agent_log is the first reference of ‘I’ and its ultimate reference is its counterpart in the real world, who is the individual John.

For this particular example, we have another account which does not involve Agent_log. We come up with a scenario in which the answering machine picks up the phone and John says (20) to the phone from some distance so that the caller can hear it. This scenario, however, does not work for (2). For it is more likely that the caller would think that ‘here’ in (2) refers to the location of John, who is distant from the answering machine, not that of the answering machine. That is, (2) would mean that John is not in the location where he is. This is not what’s meant (and is false).

6.5 Reading a Note on Behalf of its Author

Suppose that John was invited to his friend’s wedding reception, but he had an urgent business and couldn’t come. Thus, he asked Fred to read his note in the reception on his behalf. The note contains the sentence below:

(22) I wish I could be here now.

We could apply the proposed framework to this case too. The context in which Fred reads John’s note is the context of utterance/interpretation. There is Agent_log, who has the logical identity of John (and a physical identity of Fred). ‘I’ first refers to Agent_log, who will then refer to his real world counterpart, that is, the individual John. On the other hand, ‘here’ and ‘now’ are assigned their values within the context of utterance where the note is read.

6.6 Postcard

As mentioned in section 3, Michaelson (2014) proposed a theory in which the character of indexicals shift according to the context type. In the case of
postcard, the references of the indexicals are proposed as follows:

(23) Postcard: ‘I’ refers to the author, ‘here’ refers to the location of production, ‘now’ refers to the time of production. (p.528)

This works out in many cases, like ‘It is beautiful here now’ (Michaelson 2014, p.525, his (4)). However, there seem to be cases in which the reference of ‘now’ is assigned in the context of reading (decoding).

Here is an example. Suppose that John writes and sends a postcard to Fred on Wednesday. The weather forecast for Thursday is rainy, and that for Friday is sunny. John knows that Fred has a tennis match on Friday. In the postcard, John writes:

(24) Are you reading this card on Thursday? It may be raining now. But don’t worry: tomorrow will be sunny, according to the weather forecast.

In (24), ‘now’ refers to the time when the card is read, just as in the case of an answering machine, and thus it is the context of interpretation. Note, however, that when the same card is read again on Friday, the reference of ‘now’ does not change. It is the one assigned when the note was read on Thursday. Here, the preceding context (‘Are you reading this card on Thursday?’) as well as the author John’s intention play a key role.

6.7 Temporary Space vs. Stable Space

Suppose next that John lives in a dormitory. He knocks on the door of his noisy neighbor and says:

(25) I’m studying hard in my room now. Could you be a little quiet, please?

When he utters (25), John is not in his room but at his neighbor’s door. Nevertheless (25) sounds fine. Why is this? We could analyze that John’s studying in his room is a more stable context, and he is just temporarily
out of his room. In mental spaces term, John in the temporary space (at his neighbor’s door) refers to ‘his counterpart’ in the stable space (i.e. in his room). ‘I’ in (25) first refers to John as a speaker and then, by the pragmatic mapping, it refers to John studying in his room. This structure is similar to the one I proposed in step 2 of the analysis of AMP.

7. Conclusion

While face-to-face communication is a basic style of communication, there is a wider variety of communication styles, including answering machine messages and written notes. At this moment in history, new technologies are creating even more varieties, including options such as video messages and computer-synthesized utterances.

In this light, I suggest that we discard Kaplan’s (and the literature’s) empirical assumption that the agent of the context of utterance is an individual. In order to accommodate a wider range of ‘utterances’ as proper utterances, I introduced the notion of Agent_log — the agent of the context of utterance with a logical identity of an individual, lacking his/her physical identity.

The phenomenon of AMP is explained in two steps, involving both semantics and pragmatics. In the first step, the agent of utterance is identified as Agent_log and its structure is explained in terms of a context blend. In the second step, the ultimate reference of ‘I’ is identified on the basis of a pragmatic mapping from the playback context to a real world space — specifically, from Agent_log to a relevant individual. The truth of (2) is then attributed to the discrepancy between these two spaces, given that the references of ‘here’ and ‘now’ are determined in the playback context.

Kaplan’s claim about the logical truth of (1), which concerns the semantics of indexicals, is relevant only to the first step above. The only modification I made to Kaplan’s theory is an extension of the notion of the agent of the context of utterance, introducing Agent_log. With this extension, the agent of the context of utterance (Agent_log) proves to be in the context of utterance in the AMP case as well. That is, the reference of ‘I’, ‘here’, and ‘now’ are all in the context of playback and (1) holds there. With
Agent_log remaining the reference of ‘I’ within the playback context, (1) is true and (2) is false. The apparent puzzle of AMP is attributed to the second, pragmatic step. Given this, AMP is not a counterexample to Kaplan’s theory.

Furthermore, I proposed to apply Agent_log to ordinary utterances as well. That is, I argued that the semantic reference of ‘I’ is consistently Agent_log. To incorporate this idea into the theory of indexicals, I posited an additional thesis which says that ‘the agent of the context of utterance’, which appears in the character of ‘I’, means Agent_log. With this additional thesis, Kaplan’s theory of indexicals (specifically, the character of ‘I’, ‘here’ and ‘now’ and the logical truth of (1)) can successfully account for a wider variety of utterances.

References


A Mental Spaces Analysis of the Answering Machine Paradox

481-563. Oxford: Oxford University Press,