

AGENCY AND FIRST-PERSON REFERENCE

TOMIS KAPITAN
Northern Illinois University
kapitan@niu.edu

SUMMARY: In Part I of *Self-Knowing Agents*, Lucy O'Brien develops a theory of first-person reference. In what follows, I describe this account and then raise doubts about its success. Since I am not confident that I have understood her properly, I may be setting up and targeting a strawwoman, but I can only hope that what I said about the first-person will be of interest in its own right.

KEY WORDS: identification, indexical, intention, deliberation, self-consciousness

RESUMEN: En la parte I de *Self-Knowing Agents*, Lucy O'Brien expone una teoría de la referencia de primera persona. En lo que sigue describo su teoría y luego planteo dudas en torno a sus logros. Como no estoy seguro de haberla entendido correctamente, tal vez esté yo erigiendo y atacando un muñeco de paja; en todo caso, lo único que espero es que lo que se dice aquí sobre la primera persona sea de interés por sí mismo.

PALABRAS CLAVE: identificación, deícticos, intención, deliberación, autoconciencia

1. *Explaining First-Person Reference*

Typically, we speak of reference as something we do, regardless of whether we view reference as a communicative act in order to direct our audience's thought upon something, or as an act of focusing and sustaining our own attention upon that item. Expressions "refer" in a derivative sense, and then only relative to context and user's intentions (Bach 2008). First-person reference is carried out with first-person expressions for the purpose of directing attention upon ourselves. Since reference presupposes an awareness *of*, or a thinking *about*, what one is referring to, then an explanation of reference with first-person pronouns is dependent upon an account of what it is to think about oneself in a first-person way. Linguistic 'I' references are expressive of 'I' thoughts, and to explain the former is to understand their reliance upon the latter.

Throughout her book, O'Brien is as much concerned with first-person consciousness as she is with reference by means of first-personal pronouns. She does not embed her account within a more general theory of reference by means of language, and she does not attempt to explain how it is that we can think singular thoughts about anything. She focuses exclusively on the problems and proposals that arise in consideration of first-person cases, particularly upon those

generated by what she regards as four fundamental features of first-person reference:

1. First-person reference seems to be *indexical* and context-dependent. The familiar models used to explain reference by means of names and other singular concepts are inappropriate for the first person (p. 4).
2. First-person reference is *identification-free*, that is, it is expressive of an agent's self-awareness even though she has "no access to a means of identifying herself, of picking herself out as one from others." (p. 5).
3. First-person reference has an *irreducible cognitive significance*, involving a "way of referring to oneself" that differs from ways of referring to others or from other ways of referring to oneself (p. 5).
4. First-person reference is *guaranteed* in three fundamental ways, namely, a subject who uses the first-person, (a) always succeeds in referring, (b) refers reflexively, and (c) knows that she is referring to herself, that is, self-consciously refers reflexively.

If I understand O'Brien correctly, she thinks that that this last feature of first-person reference implies the well-known principle of immunity from error through misidentification relative to the first person (IEM), that is, that one cannot be mistaken that it is oneself that one refers to with the first-person pronoun (Shoemaker 1968, pp. 556–557).

Everyone understands that mere reflexive reference is not necessarily first-personal. A well-known example from the writings of Ernst Mach nicely illustrates this basic datum of self-consciousness:

Not long ago, after a trying railway journey by night, when I was very tired, I go into an omnibus, just as another man appeared at the other end. "What a shabby pedagogue that is, that has just entered", thought I. It was myself; opposite me hung a large mirror. (Mach 1959, p. 4n)

Extrapolating, we imagine that when Mach first thought,

- (1) That man is a shabby pedagogue

he did not also think,

(2) I am a shabby pedagogue

though he subsequently came to have that first-person thought. This and similar such examples are familiar means of showing that one can be aware of oneself without being self-conscious, and certainly without being aware of oneself in a first person way.

What must be added to move from “mere self-reference, that is, mere reflexive reference,” as O’Brien describes it (pp. 7, 56), to “self-conscious reference”? Both are forms of awareness of the very entity that one is, but while mere reflexive awareness is *of* an entity that happens to be oneself, self-conscious reference is more intimately tied to an organism’s sense of itself vis-à-vis its surroundings, and thus, to action. Every agent, perhaps every living organism, must have some rudimentary ability to distinguish between self and others; otherwise it could not mobilize itself to procure what it needs in order to survive (Lewis 2003, p. 280). Following O’Brien, let us use the italicized pronouns ‘*self*’, ‘*herself*’, ‘*oneself*’, ‘*him*’, etc. to speak of this specific type of reflexive awareness.

Since *self*-consciousness can be primitive and non-conceptual,¹ O’Brien recognizes that it is not going to provide enough to explain first-person reference. What more is needed to distinguish that type of *self*-consciousness that emerges in the second year of childhood and which is commonly expressed by first-person pronouns? This is the central question of Part I of O’Brien’s book.

In general terms, O’Brien’s response to this question locates the difference in the satisfaction of an epistemic condition.

The problem is an older and venerable one about first-person reference in general. In essence, the problem is that any attempt to explain

¹ As I understand it, the distinction between a *conceptual* and *non-conceptual* awareness of something is to be drawn with respect to a particular concept, property, or classification it falls under. For example, an infant might be aware of a pain, a round object, or its mother without having a concept of *pain*, *roundness*, or *mother*, and perhaps a dog can be aware of its owner or the sun without concepts of *ownership* or *the sun*. I have a conceptual awareness of this pen as a pen, though not, say, as something manufactured in Ohio, but insofar as I am aware of it then I am aware of something manufactured in Ohio. No doubt some sentient beings are aware of things they never conceptualize in any manner, e.g., their own heartbeat, the force of gravity, or life. Conceptualization typically follows upon felt contrasts and absences, and often, we fail to rise to the level of abstraction required for contrasting pervasive elements of experience.

first-person reference as ‘reflexive’ reference runs into trouble, because reflexive reference can only be first-person reference if one knows that one is referring to oneself. However, that knowledge then also needs explication. It can seem obvious however that knowing that one is referring to oneself involves referring to oneself first-personally. But if that is so it seems one cannot give a non-circular account of first-person reference. (p. 8)

Two important claims are made here that permeate her discussion in Part I. First, unlike other forms of self-awareness, first-person reference requires that the agent knows she is referring to *herself*. Second, since knowledge involves concepts, a satisfactory account of first-person reference must be a non-circular explanation that does not already assume the presence of the first-person in the explanans.

2. *Some Competing Approaches*

One might think that first-person reference is simply a matter of applying the first-person *concept* much as indexical concepts like *this*, *here*, *then*, etc. are used to conceptualize items with which we are acquainted. First-person reference is simply a matter of conceptualizing the thinker, or the thinking act, of which we are *self-conscious* and that we locate at the point of origin of our spatial or temporal fields.² Its difference from other types of reflexive awareness rests upon discriminations in terms of particularized first-person properties determined by spatial-temporal location and the creative use of the first-person concept (Castañeda 1999, chaps. 2, and 10). O’Brien does not discuss this approach in any detail, but because it relies on the identificatory powers of indexicals, it quickly falls out of her radar given the identification-free character of first-person reference.

She devotes more space to the so-called *perceptual approach* that explains first-person reference in terms of the specific manner in which we are *self-aware*, say, through quasi-perceptual introspection or some combination of exteroception and somatic proprioception. Such an approach requires that information be received *by* the subject *from* the subject through a special faculty. But O’Brien doubts whether there is any such faculty of receiving information that has exactly one object, oneself, in view. No form of exteroception or somatic proprioception can secure this, she thinks, because

²The phrase ‘point of origin’ appears in Castañeda 1977 and Evans 1982, p. 154, to designate the locale of the thinker within his or her own egocentric space. Other writers speak of the ‘locus’ of the speaker (Hintikka 1998, p. 208), the ‘zero-point’ of a locutionary act (Lyons 1995, p. 304).

of the possibility of becoming informed about someone else through their means (pp. 202–203), thereby, violating the IEM requirement.³ On the other hand, recourse to introspection is an appeal to a “mysterious” capacity of apprehending “only very enigmatic sorts of objects (Cartesian Egos and the like)” at best (pp. 6–7, 15, 97n). So, since introspection does not yield what is desired, and since perception allows for “brute error” about what we perceive (unlike knowledge of our own thoughts, pp. 96–97), then an account of *self*-knowledge first-person reference should not be modeled on the analogy with perceptual awareness.

A different strategy for accounting for first-person linguistic reference is the familiar *self-referential approach* according to which first-person reference is explained in terms of a characteristic rule, the “self-reference rule”,

SRR. ‘I’ refers to whoever produced it. (pp. 7, 49)

As stated, it is unclear what the quoted expression is intended to represent. Obviously, this rule is not talking about the first-person pronoun ‘I’ *qua* type, for the type doesn’t refer to anything. Presumably, what is being said is that a token, or an utterance, of *any* first-person pronoun, including ‘me’, ‘my’ ‘mine’ and the like, refers to whoever produced it.⁴

In any case, the core idea of this approach is that first-person reference occurs whenever a subject “comprehendingly” uses the first-person pronoun as *governed* by SRR, that is, with a grasp of

³ O’Brien takes seriously the challenge to the perceptual approach posed by the case of an amnesiac who is suspended in a state of sensory deprivation with all his senses anaesthetized, yet who is nonetheless able to think first-person thoughts. Gareth Evans, in developing his version of the perceptual approach, thinks that just as such a person can still think of the place he occupies as “here” without having any information deriving *from* that place, so too, we can think ‘I’ thoughts that are not derived from information received *from* the referent (pp. 34–35). O’Brien is not only skeptical that Evans’s perceptual model can explain how we obtain that information, but she thinks that the model cannot account for IEM (pp. 38–45). Because one can misidentify with ‘here’, e.g., thinking one is under water when one is not, and since ‘here’ can also be used as a perceptual demonstrative, then the analogy of ‘I’ with ‘here’ falls short of accommodating the immunity principle. Specifically, thoughts of the form “I am at place p” and “place p is here” which Evans suggests are the same, are not, in fact, the same, (pp. 42–43).

⁴ Because of the problematic cases like the answering machine paradox (e.g., Predelli 1998), it is necessary to give a careful statement of what it is to *produce* a token. Further adjustments needed to accommodate the uses of ‘I’ in fiction, dictation and imagination are discussed by O’Brien on pp. 50–55.

the rule and awareness of her own thoughts (p. 57). When such a subject says or thinks, “I feel warm” then he or she is referring to himself or herself in a first-person way. Anyone who exhibits a mastery of a term governed by SRR “will refer to themselves first-personally,” with the appropriate pronouns, so that “appeal” to such a rule is all that is needed to explain acts of first-person reference.

O’Brien doubts that a straightforward appeal to SRR is able to explain how it is that ‘I’ expresses *self*-consciousness. She cites David Lewis’s case of two omniscient gods, one of whom lives on the tallest mountain and throws down manna and the other of whom lives on the coldest mountain and throws down thunderbolts. Both are omniscient in the sense of knowing every proposition that is true at their world; in particular, each knows that the god who lives on the tallest mountain throws down manna and that the god who lives on the coldest mountain throws down thunderbolts. But neither knows which of the two gods he is, that is, “neither one knows whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws manna or thunderbolts” (Lewis 1980, p. 139). Suppose that both know that ‘I’ is governed by SRR and for any ‘I’ utterance either makes they know who made it. By SRR, each is guaranteed to refer to himself upon using ‘I’. So, if both gods know that the god on the tallest mountain says,

(3) I throw down manna

then both can appeal to SRR to determine the referent. In particular, the god on the tallest mountain knows that the god on the tallest mountain is referring to himself with ‘I’ in (3). Yet, this god has no guarantee that it is *himself* he is referring to. Thus, an appeal to SRR does not explain how ‘I’ is expressive of *self*-consciousness (p. 58).

Reductionist accounts which deny that there is any *self*-awareness beyond systematic reflexive awareness, are also attracted to explanations in terms of SRR (pp. 9, 59–65). According to these views, first-person reference is a matter of having attitudes the contents of which are systematically dependent upon reflexive truth-conditions, e.g., the desires and beliefs that motivate action (p. 61). But O’Brien believes that there are distinctive *self*-conscious attitudes, e.g., Oedipus’s realization that he himself is the killer of Laius, which are importantly relevant to action in a way that merely reflexive attitudes are not. Reductionist accounts cannot explain how in first-person reference the subject is referring to herself *as herself* (pp. 64–65).

3. O'Brien's Agency Account

O'Brien's explanation of first-person reference combines the subject's mastery of SRR with a basic *self*-awareness that subjects possess in virtue of being agents (p. 75). The "central idea," she writes, "is that a subject who uses 'I' in accordance with SRR, and who has agent's awareness of what she is doing, thereby simultaneously refers to herself first-personally and is able to know that she is so referring" (p. 77) Specifically, a subject who is able to refer first-personally,

A. uses the term or concept 'I' as governed by the rule that it refers to its user (*Mastery of SRR*);

B. has a capacity to know that she is referring to herself in using 'I' when she is (*Knowledge of use of 'I'*);

C. has a capacity to be aware of her own acts of thinking and uttering that does not imply a prior capacity for first-person reference (*Agent's Awareness*). (p. 77).

Because O'Brien thinks that condition B can be accounted for in terms of the more primitive conditions A and C, she feels she can defuse the circularity threatened by this epistemic requirement (see section 1 above). It is in explicating A and C that her *agency account* of first-person reference unfolds.

O'Brien describes her view as a *two-tiered strategy* that distinguishes two different levels of self-consciousness operative within *both* thought and language.⁵ The relevant distinction is between a non-conceptual *self*-awareness and a fully conceptualized *self*-awareness, with the latter explicated in terms of the former.⁶ Given that

⁵ Christopher Peacocke developed a two-tiered strategy by distinguishing between first-person reference in thought and first-person reference in language, explaining the latter in terms of former (pp. 66–71). O'Brien is unsatisfied with his approach for various reasons. For one thing, if a use of 'I' is explained in terms of thinking of oneself as the speaker of that use, then we risk violating the immunity principle since one might be mistaken about who is the speaker of an utterance one hears (p. 68). For another, when Peacocke construes the "I" concept in terms of "the one who is the thinker of these thoughts," then either "these thoughts" are tantamount to "my thoughts", in which case the first-person is presupposed, or the subject has no guarantee that the thinker of "these thoughts" is herself (p. 69). Despite these misgivings, O'Brien retains the idea that first-person pronouns are driven by first-person thoughts.

⁶ O'Brien cites Bermúdez 1998 as a precursor. Bermúdez also posited two forms of self-consciousness; full-fledged self-consciousness which requires mastery of the first-person concept and first-person pronoun, and a primitive non-conceptual self-

first-person reference is conceptual, the challenge is to explain it by recourse to non-conceptual *self*-awareness in a non-circular manner (pp. 70, 78–79, 82–83). O’Brien offers a nice comparison with the second-person pronoun to motivate her approach. It is plausible, she says, to think of a second-person reference as the joint product of (i) the speaker’s mastery of the rule that ‘you’ refers to the person addressed, and (ii) the speaker’s awareness of the person being addressed. Since one can be aware of the person being addressed in a non-second-person way, say, through a description or demonstrative, then the sense or mode through which the person is thought of *as* you is *created* by exploiting the ‘you’ concept as governed by the said rule (pp. 82–83). The first-person case is similar, though it must be explained how a subject can be aware of *herself* as the speaker and referent independently of a first-personal conception (pp. 79–82). A closer look at O’Brien’s explication of conditions A and C is necessary.

Let us begin with the mastery of SRR, which O’Brien thinks is manifested at both the level of thought and language (p. 74). What does it consist in? It does *not* require an ability to articulate SRR or to think of it as a *rule* (p. 87), but it does require that the subject

- (a) has a concept of a user of ‘I’, that is, of a *producer* of ‘I’ in thought or speech;
- (b) is capable of communication with others;
- (c) is able to *follow* SRR, not merely *conform* to SRR;
- (d) is able to intend to use the ‘I’ term or concept to refer reflexively (pp. 74–75, 87).

To have the concept of a producer, as required by (a), is to have the concepts of *thinker* and *speaker*. Presumably (c) requires that the agent has conceptualized a certain linguistic type and is able to engage in the activities of both referring with and interpreting tokens of that type. This mastery develops in early childhood when the child gains the concept of speaker and realizes that things can be

conscious that does not presuppose such mastery and which is identification-free (p. 214) and immune to error through misidentification (p. 219). This distinction allows a solution of what he calls “the paradox of self-consciousness” (p. 269). However, while Bermúdez speaks of the primitive self-consciousness as involving ‘I’ thoughts (p. 290), O’Brien demurs, correctly, in my opinion (Kapitan 2006, pp. 407–408), holding that primitive agency awareness is not a matter of thinking ‘I’ thoughts.

spoken about. Condition (d) is an elaboration upon condition (b) with respect to the first-person, and its satisfaction is critical to the agent's knowledge that she is referring to herself as herself. To execute a first-person reference, finally, requires an activation of these four capacities.

Turning to the second pillar of her account, *agency awareness*, O'Brien contends that each of us, *qua* agent, has a non-conceptual and non-perceptual awareness of our own actions that is more primitive than and, hence, independent of, first-person reference (p. 88). What is its content? In essence, it is our awareness of control over our actions, including our mental actions of judging, referring, inferring, deliberating, doubting, believing, thinking, etc. It includes the following components (pp. 115–121):

1. a direct awareness of *what* one is doing in the case of both physical and mental actions that it is not mediated by representations (it is "representation-independent") and not obtained through perceptions of what we are doing;
2. an awareness that one's doings result from one's exercise of control, and thus, from a rational assessment of possibilities.

From these we obtain a third component:

3. an awareness that one's action is one's own.

That is, since to deliberate about what to do one must be aware of possibilities as possibilities *for oneself*, then any action emerging from this deliberative process will be an action of which the agent is aware of as hers (p. 117).

In explaining how rational control over one's mental life involves agent awareness, O'Brien writes,

Suppose it is right to say that when a subject judges, doubts, or questions one thing rather than another, as a direct result of assessing what to think, she thereby has an awareness of what she is thinking. It is then plausible to suppose that, were the subject to have the appropriate concepts, this would be sufficient to entitle her immediately to self-ascribe her thoughts. (p. 117)

If the concepts in question are first-personal, then we will have secured condition B. But if the very process of exercising rational control —specifically, the assessment of possibilities for action—

already involves the sense that these are possibilities for *oneself*, how have we escaped circularity? O'Brien meets this challenge by asking us to consider whether one who is deliberating about what to believe when confronted with an inconsistency in her beliefs P , $P > Q$, and $\sim Q$ is "rationally guided" to then drop her belief that $\sim Q$ and judge that Q . O'Brien concludes that this process can occur without the subject self-ascribing her beliefs. Still, she is entitled to judge that the resulting action, namely, dropping the belief that $\sim Q$ and judging that Q , is her own.

a subject who self-ascribes an action, guided by her consideration of what to do, is entitled to take the attitude as being an action of ϕ -ing, because ϕ -ing is the action that she has practical knowledge of as a possibility, and which her consideration of what to do immediately led her to do. If we take the mode of occurrence of the mental action or activity to be given by its being the product of a process of assessment by the subject, we have reason to think that an action that occurs in that mode can stand as the reason for its own self-ascription. (pp. 119–120)

Since the "mode of occurrence" of a mental action is the product of assessment of what one should do, in this case, agent awareness (p. 115), then the mental action that occurs in that mode can stand as the reason for its own ascription (p. 120). The gap between the content "P" and "I think that P" is closed when the thought is an action that arises out of a process of agency, that is, "as possibilities that the subject has realized on the background of determining what she will do given a grasp of what she could do." In this way, a subject's act "arises as hers" (p. 120)

Two claims emerge: (i) that the agency awareness that accompanies control is independent of the capacity for first-person thought and, and (ii) that the fact that our actions are the products of our rational control rationally entitles us to self-ascribe them. Given these claims, O'Brien now offers this characterization of agency awareness:

A subject's being agent aware of her action is constituted by the action being the product of the subject's consideration of possibilities grasped as possibilities. (p. 120)

This is the Constitution Thesis for Agent's Awareness. By its means we can see that if an agent executes an act of first-person reference she has a primitive non-first-personal agency-awareness that it is *she*

who produced this token, and, through the mastery of SRR, she will know that the ‘I’ term or concept applies to *herself* (pp. 82–83, 122–124). This, then, is how condition B is met in terms of the satisfaction of conditions A and C.

Noting that it is difficult to prove a constitution claim, O’Brien thinks that the explanatory work can be achieved through a Conditional Thesis for Agent’s Awareness, namely,

If a subject acts directly on the basis of a consideration of possibilities, grasped as possibilities, then she will be agent aware of her action.
(p. 122)

This is a more modest claim to the effect that primitive agency awareness is a necessary condition of rational agency. Quite apart from the constitution claim, if we combine this thesis with the agent’s mastery of SRR we secure first-personal consciousness as we did before.⁷

4. *A Critique of O’Brien’s Account*

At first blush, it would seem that O’Brien’s account provides only a sufficient condition for first-person reference, not a necessary condition, at least not a necessary condition for having ‘I’ thoughts. She comes closer to offering a necessary condition by suggesting that while not all first-person thoughts are actions (p. 90), they can be regarded by the agent as her own insofar as she herself has *control* over them (p. 93). This allows that some ‘I’ thoughts and beliefs are passive and reflective, e.g., *I feel warm*, and are neither actions nor produced by an exercise of agency. However, the suggestion that they are under the agent’s control is dubious. Some beliefs, perceptual and otherwise, seem forced upon us by the phenomenologically given, and others, like the first-person emotions of people convulsed in anger or the self-absorbed ravings of lunatics, seem hopelessly compulsive.

⁷ One might think that only a more modest variant of this explanation is warranted. Suppose an agent acts directly from a consideration of possibilities by attempting to do an action but failing in the attempt; the intention to act and the attempt itself are still the products of deliberative process. It cannot be a necessary condition of first-person reference that the agent actually succeeds in doing the action, since the world might not cooperate. For example, I might attempt to run through a poem perfectly in my thought, yet fail, but there is definitely a sense of agency even in such a thwarted action. Agency awareness is more accurately constituted by our *attempt* to act intentionally.

Quite apart from this concern about the comprehensiveness of the account, O'Brien's optimism that she has avoided the threatened circularity of epistemic condition B is warranted only if conditions A and C can themselves be understood in a non-circular manner. I am skeptical that O'Brien has succeeded in showing this.

In attempting to secure B, she writes that "if I refer to an object as 'this' then I must know that 'this' refers to this object" (p. 81). Assuming that people, not expressions, refer, then the most reasonable interpretation of her statement is this:

- (4) If one refers to an object as 'this', then one knows that *one* is referring to the object as 'this'.

It cannot be that one identifies another person and knows that he or she is referring to the object with 'this', for this would not be a necessary condition of one's own reference with 'this'. Also, it is not enough to know merely that someone or other is referring to the object as 'this,' for this would be of little help in the effort to secure B. What is needed for that is an analogous requirement concerning use of the first-person indexical, namely,

- (5) If one refers to *oneself* through 'I', then one knows that *one* is referring to *oneself* with 'I'.

Satisfaction of condition B follows trivially from (5).

The threat of circularity is now apparent. The italicized '*one*' and '*oneself*' in (5) represent *self-consciousness*, that is, the type of reflexive awareness that lies at the base of agency awareness, not mere reflexive awareness. Since these expressions occur inside the scope of 'knows' they must represent a conceptualization by the knower of the referent. In what terms is one conceptualizing *oneself* in order to refer in first-person terms to *oneself*? If in first-person terms, then the account is plainly circular, that is, one cannot satisfy the epistemic requirement without already conceptualizing oneself in 'I' terms and, hence, satisfying that requirement cannot be the explanation for how first-person consciousness differs from basic *self-consciousness*. If not in first-person terms, then how else does this self-conceptualizing occur?

So, if in order to satisfy condition B, one must know that *one* is referring with 'I' to *oneself* when one does so, and if this identification of *oneself* is already first-personal, then we are relying on first-person consciousness —upon 'I' thoughts— to explain first-person reference.

This is not a bad thing if we are simply explaining speaker's reference with first-person pronouns. But O'Brien has the more ambitious project of explaining first-person *thoughts* as well, and, in that case, the threat of circularity is not easily dismissed.

Perhaps the threat can be defused if we fully appreciate how agency awareness and a mastery of SRR generate first-person thoughts. Let us turn, once again, to these conditions. In a revealing passage, O'Brien writes that even if we accept the constitution thesis as accurately describing agency awareness, there is still more explanatory work to be done since the account

is going to have to draw on yet more primitive notions of awareness, in particular, the subject's awareness of the possibilities available, and the awareness of them as possibilities. If rational assessment brings about a form of self-awareness when we are aware of ways we might act, we are still going to have to explain that prior awareness of such templates for action. (p. 122)

It is interesting that O'Brien does not pursue the issue, because if she had then she might have encountered circularity for yet another source. Consider; according to the constitution thesis, the action that an agent recognizes as one's own through agency awareness is presumably awareness of one's performing of an action at a time. It is recognized as such because it is a product of a rational assessment of possibilities. What sort of thing is a *possibility* that can be the content of a rational assessment? O'Brien does not tell us. The most charitable interpretation is that she is talking about a possible action. But what is that? Obviously, it is not the *event* consisting of one's actually doing something at a particular time. More plausibly, we are talking about a way one might act, that is, about an action-type, as suggested by the schematic language at the top of page 120.

Unfortunately, this is not enough. I often deliberate about *whether* to go to my office by bicycle or to go by car, recognizing both modes of travel to be possible for me. I usually do this shortly before leaving the house, and what I think about doing is typically something I can accomplish within the next 15 minutes, regardless of which mode I select. One might suppose that the "possibilities" I am assessing can be portrayed through the gerundial forms,

going to my office by bicycle

going to my office by car

both of which express action-types. However, neither of these grammatical forms juxtaposes with ‘whether’ to accurately depict the contents of my deliberation. A better pair of candidates are the infinitive forms,

to go to my office by bicycle

to go to my office by car

respectively. These grammatical forms represent ways of acting as something *to* be done rather than as something considered as *being* done, and to that extent, are more accurate portrayals of what is deliberated about.⁸

Moving to the infinitive forms brings us closer to what is being assessed, but we still fall short. What I am deliberating about is what to do *in* the very near future, in which case it is more accurate to portray the possibilities by incorporating the time factor expressed by ‘within the next 15 minutes’, thus,

to go to my office by bicycle within the next 15 minutes

to go to my office by car within the next 15 minutes.

Yet, still more is needed if we are to locate the proper bearer of the modality. After all, it is not these temporally particularized ways of behaving that are, strictly speaking, possible for me; rather, what is thought to be possible is that *I* perform them, not that some other person, whether identified or not, performs them. If so, then the agent of the considered action must be represented as well, that is, myself, as in,

myself to go to my office by bicycle within the next 15 minutes

myself to go to my office by car within the next 15 minutes.

Let us say that each of these represents a *course-of-action*, each of which has the form “X, to do A at time t” (Castañeda 1975).

Now we face a critical decision. If deliberation involves a rational assessment of courses-of-action, each of which is a way of acting at a particular time for oneself, *and* if to deliberate is to conceptualize

⁸ See Castañeda 1975, who distinguish practical thought contents (*practitions*) from contemplative thought contents (*propositions*), with the former represented by the infinitive form. He argued persuasively that the immediate contents of deliberation and intention are *practitions* rather than *propositions*.

those possibilities, then we have *self*-reference creeping back within attitudinal scope, namely, into constructions of the form ‘deliberates whether’ and ‘rationally assesses’. If this *self*-reference is first-personal then, once again, we must abandon the attempt to explain first-person in terms of agency awareness. The only way to escape this circularity is to insist that the reflexive awareness involved in the consideration of possibilities is nothing more than the basic type of *self*-consciousness that accompanies all agency. It sounds strange to think that this is the reflexive awareness involved in *rationally assessing* one’s possibilities, since ‘rationally assessing’ connotes a fairly sophisticated mode of consciousness and articulated awareness of *what* one is assessing.

However, even if this move succeeds in avoiding circularity, the acknowledgement that self-consciousness already permeates the contents of deliberation renders the constitution thesis trivial. Without doubt, an agent whose performance is a result of rational assessment will regard that performance as *her own* since the considered possibilities are already regarded by her as *her own*. Thus, insofar as it explains anything, the constitution thesis does not reveal how it is that I am conscious that an action is identified by me as *mine*. At best, it shows only which one among the various courses-of-action that I am already aware of as possibilities *for me* I am actually performing. Similarly, the conditional thesis carries no explanatory force since the self-awareness spoken of in the consequent is already required in any satisfaction of the antecedent.

If we escape circularity in the explication of agency awareness, the appeal to a mastery of SRR is not so fortunate. In understanding how indexical reference works, it is vital to distinguish a producer’s *execution* of an indexical utterance from a hearer’s *interpretation* of that utterance (Castañeda 1983, Kapitan 2001). No doubt, mastery of SRR is required to interpret a speaker’s first-person pronouns and determine the speaker’s referent. But the hearer’s interpretive process does not occasion a first-person ‘I’ thought about *that* referent, save in the case of talking to oneself. Consider, if I say to you,

(6) I will bring you a beer if you stay for a while

then you will rely on your mastery of SRR to understand what I am saying. You won’t think an ‘I’ thought about *me* in the process, but, very likely, a third person or demonstrative thought better expressed by,

(7) He’ll bring me a beer if I stay for a while.

By contrast, I, the producer, already have an ‘I’ thought about myself that is partly responsible for my utterance of (6), and it is apparent that I do not think of myself demonstratively in the way you do. Nor need I think of myself as a producer of a first-person reference, that is a thinker or speaker, e.g., as,

(8) The speaker of (6) will bring you a beer if you stay a while.

for such a thought is not first-personal at all.

The example gives rise to two important questions:

- Q1. Does SRR govern a speaker’s use of ‘I’?
 Q2. Does SRR govern a speaker’s use of the first-person concept, thus, of ‘I’ thoughts?

The answer to Q1 is affirmative assuming that the speaker knows something about how the hearer will process the perceived ‘I’ token. *Qua* speaker, he intends that the hearer will interpret ‘I’ as referring to himself *qua* producer of that token. Moreover, he intends that the hearer understand that he is referring to himself in a first-person way, thus, that the speaker is thinking ‘I’ thoughts. In these two ways, then, SRR governs a speaker’s use of ‘I’.

However, the speaker’s own ‘I’ thoughts are not generated in this fashion. That is, the speaker/thinker does not *come to have* an ‘I’ thought by exercising the capacities constitutive of the mastery of SRR. As indicated, that rule is suitable only for an *intention* to communicate via ‘I’ and for an *interpretation* of an already existing ‘I’ utterance or an ‘I’ thought (if it is even sensible to speak of interpreting an ‘I’ thought). In general, just as we do not causally explain production in terms of its products, we do not explain how a thinker comes to make a first-person reference in terms of interpreting his or her own first-person references. One does not identify ‘I’ tokens, ‘I’ utterances, or ‘I’ thoughts and then exercise the SRR capacities in order to *think* in first-person terms. Patently, an interpretive consumption of such elements cannot precede their production. The occurrence of particular ‘I’ thoughts within the speaker are *causes* of his or her ‘I’ utterances, not the effects of those utterances. If so, then the exercise of the SRR capacities by the thinker is not even a necessary condition of first-person reference, and it is certainly not what explains the thinking of first-person thoughts. The answer to Q2, accordingly, is negative.

Here then is the circularity. A speaker who produces first-person utterances intends that the hearer understands that the speaker is referring to himself or herself in a first-person way. Obviously, this intention embodies first-person thinking, and since the intention produces the utterance, then the speaker does not have to interpret his own utterance through SRR in order to think in a first-person way. Mach did not have to utter a sentence like

(2) I am a shabby pedagogue

in order to think the first-person thought that he himself is a shabby pedagogue. Nor did he first think the first-person thought and *then* understand that it is himself that he is thinking about *because* he understood who produced that 'I' token in thought. He could not be aware of that token without producing it, and to produce it is already to have been aware of oneself via first-person concept. Similarly, my uttering (6) is brought about by my thinking a first-person thought, but my thinking that thought is occasioned by my determining who it is that produced an 'I' token, either in speech or in thought. Thus, an appeal to SRR is not the key to understanding first-person thinking even if the capacity to use first-person pronouns as governed by SRR is a necessary condition of first-person communication.

In sum, O'Brien is effectively placing the cart before the horse in her account of first-person reference. A mastery of SRR is relevant to understanding reference only if a subject applies the constitutive capacities to already constituted first-person utterances or thoughts. First-person utterances are the products of first-person thoughts, and the latter exist only if the speaker already thinks in first-person terms. But since to think in first-person terms requires the possession and *use* of first-person concepts, and is not the *result* of exercising our SRR capacities, it follows we do not *explain* first-person reference by combining *self*-consciousness with a mastery of SRR.

5. Concluding Remarks: First-Person Identification

There is a simpler approach to first-person reference that, like O'Brien's, utilizes a primitive form of *self*-awareness in the explanans but that invokes indexical identification in terms of an *I* concept. I developed such an account in Kapitan 2006, and it departs from O'Brien's in two fundamental ways. First, it sacrifices the claim that first-person reference is identification free. Second, since indexical identification involves *locating* the referent, it agrees with one aspect of the perceptual approach, namely, that in first-person reference a

subject must be aware of *something* as being located at the point of origin of his or her own perspective, if not the referent then some part or aspect of the referent.⁹

The account readily accommodates the other noted features of indexical reference (see section 1 above). Each indexical concept comes with its own constraints, e.g., *then*, *now*, and *before*, are invariably used to conceptualize events, processes, and intervals. Similarly, because the concept of *I* can only be applied to a *self*-aware experiencer located at the point of origin of a perspective (Kapitan 2006, p. 396), the IEM condition is secured. Once we reject the identification-free feature of first-person reference, and insist that we can identify what is already there in experience via a particularized first-person property, *viz.*, being an *I* located at the point of origin of an experience (which we represent as *here* and *now*) then we have no difficulty explaining the difference in Oedipus's behavior upon realizing that he himself killed Laius, or how it is that two omniscient gods can tell themselves apart. There is then no need to explain first-person reference in terms of the subject's knowledge that she is referring to *herself*. Such knowledge is, at best, a consequence of first-person reference, not its source.

REFERENCES

- Anscombe, G.E.M., 1975, "The First Person", in S. Guttenplan (ed.), *Mind and Language*, Oxford University Press, Oxford, pp. 45–64.
 Bach, Kent, 2008, "On Referring and Not Referring", in J.K. Gundel and H. Hedberg (eds.), Oxford University Press, Oxford.

⁹ In Kapitan 2001, I presented an account of indexicals wherein indexical identification occurs through application of an indexical concept to something one is immediately aware of within one's perspective. This something is an *index*, *viz.*, the "thing picked out by the deictic component of an indexical" (Nunberg 1993, p. 19) or what one "latches onto" in an instance of indexical thinking (Anscombe 1975, p. 54). In the case of an immediate indexical reference, this index is classified in terms of an indexical concept, thus, *this is darker than that*, pointing to two colored disks in one's visual field. If the reference is deferred, then the referent is classified indexically by relation to the index, as in *this city is to the north of that one*, pointing to a dot on a map. In Kapitan 2006, I followed Whitehead 1958, p. 224, and Castañeda 1999, pp. 244, 263, in arguing that in the case of first person reference, the index is an event of experiencing, that is, a short-lived process of uniting diverse factors into the unity of one perspective and located at the point of origin of that perspective (Kapitan 2006, p. 398). First-person references can be either immediate, with the identified *I* being identical to the index, or deferred, with the referent identified in terms of such an index, say, a past stage of a person or the entire enduring organism.

- Bermúdez, José, 1998, *The Paradox of Self-Consciousness*, MIT Press, Cambridge, Mass.
- Castañeda, Hector-Neri, 1999, *The Phenomeno-Logic of the I: Essays on Self-Consciousness*, ed. James G. Hart and Tomis Kapitan, Indiana University Press, Bloomington.
- , 1983, “Reply to John Perry: Meaning, Belief, and Reference”, in J. Tomberlin (ed.), *Agent, Language, and the Structure of the World*, Hackett, Indianapolis, pp. 313–328.
- , 1977, “Perception, Belief, and the Structure of Physical Objects and Consciousness”, *Synthese*, vol. 35, pp. 285–351.
- , 1975, *Thinking and Doing*, Reidel, Dordrecht.
- Evans, Gareth, 1982, *The Varieties of Reference*, Oxford University Press, Oxford.
- Filmore, Charles, 1997, *Lectures on Deixis*, CSLI Publications, Stanford.
- Hintikka, Jaakko, 1998, “Perspectival Identification, Demonstratives, and Small Worlds”, *Synthese*, vol. 114, pp. 203–232.
- Kapitan, Tomis, 2006, “Indexicality and Self-Awareness”, in Uriah Kriegel and Ken Williford (eds.), *Consciousness and Self Reference*, MIT Press, Cambridge, pp. 379–408.
- , 2001, “Indexical Identification: A Perspectival Account”, *Philosophical Psychology*, vol. 14, no. 3, pp. 293–312.
- Lewis, Michael, 2003, “The Development of Self-Consciousness”, in Roessler and Eilan 2003, pp. 275–295.
- Lyons, John, 1995, *Linguistic Semantics*, Cambridge University Press, Cambridge.
- Mach, Ernst, 1959, *The Analysis of Sensations*, Dover, Nueva York.
- Nunberg, Geoffrey, 1993, “Indexicality and Deixis”, *Linguistics and Philosophy*, vol. 16, pp. 1–43.
- O’Brien, Lucy, 2007, *Self-Knowing Agents*, Oxford University Press, Oxford.
- , 2003, “On Knowing One’s Own Actions”, in Roessler and Eilan 2003, pp. 358–382.
- Peacocke, Christopher, 2001, “First-Person Reference, Representational Independence, and Self-Knowledge”, in A. Brook and R. DeVidi (eds.), *Self-Reference and Self-Awareness*, John-Benjamins, Amsterdam.
- Predelli, Stefano, 1998, “Utterance, Interpretation, and the Logic of Indexicals”, *Mind and Language*, vol. 13, pp. 400–414.
- Roessler, Johannes and Naomi Eilan (eds.), 2003, *Agency and Self-Awareness*, Oxford University Press, Oxford.
- Shoemaker, Sydney, 1968, “Self-Reference and Self-Awareness”, *The Journal of Philosophy*, vol. 65, pp. 555–567.

Received: March 2, 2009; revised: January 27, 2011; accepted: January 28, 2011.