Mats Andrén

Centre for Cognitive Semiotics, Lund University

The Social World Within Reach:
Intersubjective Manifestations of Action Completion

This is a study of the intersubjective recognizability of the ‘proper’ accomplishment of children’s actions: in particular, how the status of actions as properly completed is often actively made recognizable through speech and various modulations of bodily movement. In addition to analyzing how children do this in a number of cases, I argue that these manifestations of action completion are often strongly dependent on typified conventional knowledge, and that conventionality on the side of the signified is a neglected issue in gesture research. The data consists of video recordings of four Swedish children between 24–30 months of age who interact with their parents at home. The analysis is framed in ideas about intersubjectivity and action drawn from Alfred Schutz and Adam Kendon in particular, but also others. These theoretical syntheses are a substantial part of the contribution of this paper.

Keywords: gesture, child-parent interaction, social activity, intersubjectivity, materiality

INTRODUCTION

Within the literature on gesture, the relation between empty-handed gesture and acts that involve objects has generally been treated as synonymous with the distinction between instrumental and communicative action – sometimes for theoretical and sometimes for methodological reasons (Andrén 2010). Thus, communicative acts that involve handling of actual objects have received very little analytical attention, even though such acts occur frequently and in many forms. I will analyze one particular aspect of such actions in detail: namely, cases where the status of an action as accomplished and completed is not merely a matter of being recognizable and publicly available for those who happen to be attending; rather, it is made recognizable in certain ways. Children learn not only a repertoire of actions but also typified means for making acts visible to others. The analysis here is focused precisely on uncovering some of the aspects
of how children do these things and how that relates to the intersubjective appearances and treatment of their acts more generally. Particular attention is paid to the role of conventionality in these processes. The study is based on monthly video recordings of four Swedish children between 24-30 months of age as they interact relatively freely with their parents, at home, in several different kinds of activities. The analysis is framed in ideas about the role of typified action in intersubjectivity, drawn in particular from Alfred Schutz; and ideas about the visible features of bodily acts that appear communicative to us, drawn in particular from Adam Kendon, complemented by concepts from Aleksei N. Leontiev, Per Linell, Erving Goffman, and gestalt psychology. These theoretical syntheses, including extensions to the originally formulated concepts, are among the contributions of this article.

The present study is part of a larger effort to understand better the expressive potentials of bodily actions that involve the handling of objects. Andrén (2010) provides a complementary treatment of the expressive properties of other phases of object handling, such as when the hand approaches an object, grabs or otherwise touches it, picks it up, does something with it while it is in possession, puts it down again, and withdraws from it. All of these phases of action may take on expressive properties in ways that are particular to the ‘ecology’ of the particular phase. The theme of the present study is to elaborate on the character of the last of these phases: namely, the phase when there is some form of withdrawal from the object, which has just been used somehow.

The article is organized as follows. First comes a discussion of Alfred Schutz’s ideas relating to intersubjectivity, action, and typification. Next, there is another theoretically oriented section, which deals with the properties of bodily actions that appear communicative – focused on the gestalt-like appearances of gesture and the interplay between parts and wholes in bodily expression. Then there is an analysis of children’s communicative acts involving objects and various manifestations of action completion in relation to these acts. Finally, a concluding discussion concerns the nature of the empirical findings and their theoretical implications.

**INTERSUBJECTIVITY AND THE WORLD WITHIN REACH**

Alfred Schutz was concerned with establishing a phenomenology of the natural attitude in the activities of everyday life. As part of this endeavor, he analysed *the world within reach* (Schutz 1945) along with its intersubjective properties. ‘The world within reach’ is the world as experienced through the senses as well as kinaesthetic, locomotive, and operative movements that ‘gear out into the world’. Sometimes Schutz

---

1 Three of the children are part of the Strömqvist-Richthoff corpus. See Richthoff (2000) for more details. I collected the data for the fourth mother-child dyad. All communicative acts (i.e., utterances and communicative bodily actions) involving the children has been analyzed and coded in ten hours of the data. This article deals only with a few examples taken from the corpus. The coding system is not an explicit part of the analysis presented here and is therefore not described.

2 ‘Within reach’ is a translation from the German word *erreichbar*; it seems to be the standard translation of Schutz’s term.
uses other expressions, such as *zone of operation* (Schutz & Luckmann 1973) or *manipulatory sphere* (Schutz 1955), referring to the more specific parts of the world susceptible to direct action. His analysis of this part of the *life-world* was inspired by George Herbert Mead (1932, 1938), who distinguished between the *manipulatory area* and more distal zones. Schutz regarded the manipulatory area as being of utmost importance, describing it as ‘the kernel of the reality of the life-world’ (Schutz & Luckmann 1973: 42; *see also* Schutz 1945: 546). Although terms such as ‘zone’, ‘sphere’, and ‘area’ may suggest otherwise, this zone is not to be understood merely as a *spatial* zone. It is endowed with the *temporal* experiential character of the present, including various relations to past experiences and future possibilities, in terms such as ‘the world within potential reach’ and ‘the world within restorable reach’, and through temporally situated motives such as doing things ‘in order to’ (i.e., to achieve a future result) and ‘because of’ (i.e., motivated by some previous state of affairs.) The concept of the *world within reach* is framed in a solid spatiotemporal and experiential framework.

The most crucial aspect in the analysis of the *zone of operation* is that it is an analysis not only of the relation between the subject and the world, but also of action in the world, as a publicly available and *intersubjective* phenomenon. The *other* and the *we*-relation are never forgotten. The importance of triadic relationships – between the subject, the *other*, and objects – are recognized. Schutz paid close attention to the partly overlapping, partly disjoint, character of participants’ experiences and perspectives in face-to-face encounters: for example, carefully working out the consequential nature of the distinction between *what is here for me is there for you* and vice versa. Questions of intersubjectivity constitute the very heart of Schutz’s theoretical enterprise: intersubjectivity provides the methodological foundations to the social sciences (Schutz 1954). Indeed, whenever embodied approaches to human thought and action reduce the *other* to a mere part of the environment, without an explicit account of the intersubjective conditions of *social* action, a significant omission has been made – since the publicly visible and shared character of worldly engagement has very real consequences for the organization of action. Examples of more or less ‘embodied’ accounts – of the subject in its environment – that lack a developed notion of action in the social sense include Johnson’s (1987) ‘body in the mind’, the ‘ecological’ framework of Gibson (1979), von Uexküll’s (1926) ‘functional circles’, and the action-perception account of Noë (2004).

---

3 This is, to some extent, a matter of degree: both in terms of the continuous nature of distance itself and the way various artefacts and tools extend our reach in various ways. Schutz did not want to imply that there is a fixed boundary between the manipulatory sphere and the spheres of potential action.

4 Kendon (1990: 211) presents a similar notion in a discussion of spatial organization in communication. When an individual’s *transactional segment* (the space into which a participant looks and speaks and into which he reaches to handle objects) overlaps with another individual’s transactional segment in a *joint transactional segment* (an *o-space*), one can speak of a jointly regulated *F-formation*: a stable spatial arrangement of a focused encounter.

5 See Gallagher (2008b) for a critique of Noë on this point, although on different grounds than discussed in this article.
In the literature that does stress the importance of the social aspects of action and intersubjectivity, one can, ideally speaking, discern two poles regarding how the problem of intersubjectivity is construed. The benefit of Schutz’s account is that it provides a balanced antidote to these two extremes. On the one hand, one has approaches that start out with the solipsistic assumption of ‘no access’: minds first exist in separation, and some sort of mental bridge is required to close the gap. In such accounts, it is typically minds or brains, rather than persons or bodies, that make intersubjectivity possible; emphasis lies on domain-general inferential mechanisms: *theory of mind* (Baron-Cohen 1995) and, to some extent, the relatively detached *shared intentionality infrastructure* of Tomasello (2008). These accounts are often insensitive to the ways in which most cognitive processes rely, in part, on actual and diverse experiences and knowledge (‘the stock of knowledge at hand’: Schutz 1953: 4), which have their basis in a previous sedimented history of engagement in various activities. According to the extreme version of the ‘no access’ position, interpersonal understanding is solely possible solely by means of powerful inferential processing of a domain-general kind. As a result, the problem of interpersonal understanding tends to be viewed as relatively ‘content-free’ (highly generalized), detached from more specific conventionalized practices and rationales, and principally separate from particular ways of engaging with others. Note that Schutz speaks, not of interpersonal *understanding* but of ‘knowledge of other minds’ (my italics; see Schutz 1954: 265); intersubjectivity is a practical matter of a heterogeneous set of everyday activities rather than a general and homogeneous sort of problem. This way of viewing the problem has a distinctly more concrete touch to it, bringing to the forefront the fact that interpersonal understanding is not just one thing but a diverse set of problems and phenomena (cf. Zahavi 2010: 299).

On the other hand, one has approaches that start out with an assumption of access to the other. According to this position, it is a mistake to think that persons are separate from each other in the first place; they are, in fact, physically and experientially available to each other, through emotional expression and the like. These approaches generally emphasize some form of *direct perception* of the other and are mainly found within the phenomenological tradition. By way of example, Schutz (1948: 191) writes that, for Sartre, ‘expressive gestures in particular, do not indicate a hidden affect lived through by any psyche. The frowning brows, the clenched fists, etc., do not indicate the Other’s wrath; they *are* his wrath.’ Max Scheler (1954) defends a similar view. This way of formulating the problem avoids some of the difficulties associated with the previous approach, but it also comes with its own difficulties. Such views sometimes come dangerously close to accounting for intersubjectivity merely by saying that it is not a problem in the first place. That may true to some extent – from the viewpoint of participants in routine activities – but it is not necessarily productive from a scientific point of view. It also comes close to a behaviorist position, as Schutz (1948) notes, such that the mental life of a subject at any given moment is reduced to what is

---

6 See Gallagher (2004) for a related argument based on the comparison of inferential and direct perception accounts of action understanding, with relation to Aaron Gurwitsch rather than Schutz.
visible from the surface of the person’s body. The difficulty sometimes becomes confused with the question of Cartesianism. Just because one does not want to separate mind from body in a Cartesian way does not mean one wants to reduce mind to that which is visible on the surface of the body in any given moment – neither from a scientific point of view nor as a description of how participants orient to each other.

The way Schutz construes the problem of intersubjectivity, however, one is neither in a position of ‘no access’ to the other, nor does one have direct access to all aspects of the other’s mindful life. What two people have between them is the social world of everyday life in the form of activities that involve the other from the outset. Schutz (1955: 173) states that ‘fully successful communication’ is unattainable. There should be no illusions of unproblematic ‘direct’ access to the other. ‘The commonsense praxis of everyday life, however, solves this problem to such an extent that for nearly all good and useful purposes we can establish communication with our fellow-men and come to terms with them’ (1955:173). Social order – in this pragmatic and practical sense – is only possible if the communicative process is augmented by various typifications, routinized standardizations, and abstractions. Thus, he locates the problem of intersubjectivity right in the middle of the spatiotemporal, personal, and mundane contingencies of coping with everyday life. He also assigns a key role to conventionalized and typified ways of acting and interpreting action – typical motives for typical actions – in this process. Intersubjectivity, he says, ‘is not a problem of constitution which can be solved within the transcendental sphere, but is rather a datum of the life-world’ (Schutz 1966: 82).

When people act, they orient to more or less standardized, tried-out-before, typified ways of acting, rather than being strictly governed by them (cf. Seedhouse 2007). They do so because it works, and because others can understand them, to a sufficient degree for current practical purposes, when they do so. To a large extent, they also understand the actions of others in terms of such standardized motives and rationalities (Schutz 1953). Through the socially shared character of this sort of action understanding, people are able to achieve and maintain a social order sufficient for the pragmatic purposes at hand.

This means that, if one can provide a firm account of these more or less typified processes of social activity, one has come a long way toward understanding intersubjectivity itself. (This is not to say that all

---

7 See De Jaegher (2008) for a recent, phenomenologically inspired account that emphasizes the importance of joint action for interpersonal understanding.

8 It has been disputed whether Schutz’s critique of Husserl’s transcendental account of intersubjectivity amounts to a rejection of the very possibility of a transcendental analysis of intersubjectivity, or whether it merely points to limitations in the particular account provided by Husserl (Zahavi 2001: 23-24). In either case, Schutz’s own account of intersubjectivity provides a useful framework for the analysis presented in this paper. This does not mean that other approaches to intersubjectivity must be rejected.

9 This way of thinking was picked up by Harold Garfinkel, the founder of ethnomethodology, and made into a central point of departure for this school. According to ethnomethodological thinking, social life is organized with respect to being visibly-rational-and-reportable-for-all-practical-purposes (Garfinkel 1967: vii). Garfinkel’s (1952) PhD thesis provided an account of social order based, to a large extent, on Schutz’s ideas, although Garfinkel was also partly critical towards Schutz (and Mead) (Garfinkel 2006 [1948]).
aspects of the problem of intersubjectivity are reducible to visible aspects of action.) As Heritage (1984: 49) pointed out in a discussion on Schutz, this typified interpersonal understanding – really an extended interpretation of Max Weber’s (1947 [1922]) concept of Verstehen – is not a matter of actually having access to others’ experiences. At the same time, it does not mean that interpersonal understanding is arbitrary, lacking substance or orderliness. According to Schutz, social constructions – such as the various typified interpretations and procedures – are rather what provides the very possibility of social order, albeit not necessarily a perfect one (whatever that would be). This take on order and mutual understanding in social life stands in contrast to the sometimes one-sided theoretical treatment of conventionality and culture whenever the discussion of social phenomena is limited to terms like ‘variability’, ‘arbitrariness’, and ‘cultural differences’ (the curse of Babel), rather than focusing on the power of the social processes of typification for the creation and maintenance of coherence and order within groups – which is, I suggest, the primary function of conventionality. Schutz (1954: 264) explicitly rejects a relativist standpoint, describing the typifications that underlie Verstehen as:

…The particular experiential form in which common-sense thinking takes cognizance of the social cultural world. It has nothing to do with introspection, it is a result of processes of learning or acculturation in the same way as the common-sense experience of the so-called natural world. Verstehen is, moreover, by no means a private affair of the observer which cannot be controlled by the experiences of other observers.

The further away from the manipulatory sphere one gets, and the less of an actual face-to-face encounter an interpersonal meeting is (telephone calls, conversations through letters, etc), the stronger the dependence on indirect processes based solely on established typifications, such as conventionalized language and standardized activities. These include the ways in which people typify their own conduct in institutional settings, in the sense of ‘acting like the typical sender of a letter’ at the post office (Schutz & Luckmann 1973), and the like. That said, it is in the we-relationship of the face-to-face encounter and the world within reach that children first find themselves, when they are born into the world (Schutz 1967 [1932], 1966: 80). Note, though, that even in the case of face-to-face encounters, Schutz did not accept uncritically the idea of perception of the other as ‘direct’ – even from a phenomenological perspective (Schutz & Luckmann 1973: 63-64):

Although we speak of the ‘immediate’ experience of a fellow-man, this experience is internally, also in the precise meaning of the word, ‘mediated.’ I grasp my fellow-man’s flow of lived experiences only ‘mediately,’ in that I explicate his movements, his expression, his communications as indications of the subjectively meaningful experiences of an alter ego. But among all my experiences of the other I, what is mediated least is the encounter of the fellow-man in the simultaneity of the we-relation. Thus we will continue to speak, even though it is not completely accurate, of an immediate experience of the fellow-man.

10 Mead (1926), Bruner (1990), and Linell (2009: ch. 20) argue in similar ways for a take on social constructions as ‘real’: i.e., in general accord with the so-called Thomas theorem (Thomas 1928), which states that ‘if men define situations as real, they are real in their consequences’ – but not in a way that entails relativism in its stronger forms. Rather, their positions are compatible with a form of realism.
Unless one allows for some kind of indirectness, it is hard to explain what it is that adults can do that infants cannot, or what humans can do but apes cannot, or how communication could ever go wrong. So in Schutz’s view, and my own, the question of interpersonal understanding is not reducible to public activity and conduct. Nonetheless, the public character of action is of primary importance in any account of intersubjectivity. The way Schutz formulates the problem productively locates a phenomenon that is readily available for analysis each time human beings engage with each other; which is not to say that other ways of studying intersubjectivity must be excluded. Schutz’s formulation is also insightful because of the way it realizes the key role of convention and typification in the process.

MANUAL EXCURSIONS AND ACTION GESTALTS

This brings me to the question of what it is that make people see certain movements as social or communicative. Schutz and Luckmann (1973: 52) state that it is ‘…as Husserl has shown, a universal principle of consciousness that in my conscious acts I “live” attentive to their intentional Objects, not to the acts themselves.’ In the case of gesture, this means that one generally does not attend to the hands or the movements of the hands as such; one rather sees ‘through them’ to some sort of content or referent (cf. Kendon 2004: 358, Streeck 2009: 139). This is the phenomenological sense in which it could be said that one perceives directly the intentional objects of acts performed by others.\(^\text{11}\) In the case of (empty-handed) gesture, Kendon (1978) made gesture researchers thoroughly aware that only some parts of the manual movements required to realize a gesture are perceived as ‘the gesture’ – standing out as an expressive figure against ground – even though generally there is movement going on both before and after. Kendon has called such communicatively articulated segments of bodily movements the stroke(s) (Kendon 1980), defining them as those parts of the manual excursion ‘…in which the movement dynamics of “effort” and “shape” are manifested with greatest clarity’ (Kendon 2004: 112). The movement phase coming before the stroke he calls the preparation phase. After the stroke, another preparation phase may follow (in cases where several gestures are performed in sequence), or there may be a retraction phase\(^\text{12}\), consisting of the return of the articulating part to its rest position (Kendon 1980: 212). Although not strictly necessary, gesture retractions have a strong tendency to end where the initial preparation phase began; this is the reason for the use of the term ‘excursions’. One may find, at various junctures during such movement excursions, holds, where the hand (the most common articulator of gestures) momentarily pauses in the course of a not-yet finished movement excursion.

Kendon (2004: 14) argues that precisely these dynamic features of gestural movements, captured in the analytic vocabulary of gesture phases as just described, make observers see some movements as

---

\(^{11}\) Of course, this ‘directness’ is from the first-person perspective. One proponent (Gallagher 2008a: 553) of a direct-perception account of others’ intentions puts it like this: ‘I also admit that one can identify various articulations at the subpersonal level and find nothing at that level which can be labeled “direct” in regard to perception.’

\(^{12}\) The term ‘retraction’ is used by most gesture researchers; however, Kendon prefers the term ‘recovery’.
intentionally communicative. He writes that ‘deliberate expressiveness is manifest’ (2004: 14). Due to this we perceive ‘the quality of the action as intentional (not the specific intention, necessarily)’ (2004: 14), without having to infer any intention explicitly. As I have suggested elsewhere (Andrén 2010), the mere fact that many gestural excursions are performed with empty hands may, itself, be one such manifest feature of ‘deliberate expressiveness’, perhaps even a particularly important one. Empty-handed excursion-like movements will have a very strong tendency to appear (a) volitional, due to their excursion-like character; and (b) performed for communicative rather than practical purposes, due to the absence of any object. This should not be taken as grounds for treating the absence of objects as a necessary feature in definitions of gesture, even though this is precisely what many gesture researchers do. Defining away communicative acts that involve objects does not make them go away. Kendon does not himself commit the mistake of excluding communicative acts that involve objects from the research agenda. That said, he does not, in practice, analyze this class of actions, and therefore it remains unclear from his account of gesture phases how and if they might apply to communicative acts that do involve objects.

Streeck (2009: 23) writes that ‘…in the context of conversation, it is usually not difficult at all to identify gesture units and their boundaries. In contexts of work, however, gestural communication may consist in nothing more than a repetition or a slight embellishment or exaggeration of an instrumental act.’ However, it should be emphasized that communicative acts that involve objects are not necessarily constituted only of such ‘slight’ embellishments. In many cases, they are clearly articulated. As I will show later on in the analysis, Kendon’s vocabulary of empty-handed gesture phases remains useful for the description of many communicative acts that involve objects – although it requires some elaboration. Acts that involve objects often have the character of excursions, similar to empty-handed gestures. They likewise often have a similar phase-like structure in terms of preparation, stroke (the ‘actual’ act), and retraction.\textsuperscript{13}

The analysis presented in Section 4 should be viewed as part of a larger project of understanding the spatiotemporal and intersubjective coordination of action in Schutz’s (and Mead’s) ‘manipulatory area’ – taking Kendon’s analysis as a point of departure for a more detailed analysis of communicative bodily movements, while extending it to incorporate the specific conditions of communicative acts that involve the manipulation of objects. What frequently happens in social settings where objects are used is that the object-directed acts enter into the \textit{main attentional track} of the activity (cf. Kendon 1978, Goffman 1974). Under such circumstances, the common division between the practical and the communicative has a tendency to dissolve, and many acts appear to be hybrids of these two aspects of action.

\textsuperscript{13} Sacks and Schegloff (2002) formulated the concept of ‘home position’, more or less equivalent to Kendon’s concept of ‘rest position’, around the same time as Kendon in the ’70s, although they did not publish it at the time. They write that, even though they first noted the phenomenon in gestures – i.e., gestural excursions often end where they begin – it applies to (some) actions of a more practical kind, too (p. 141).
One key feature of Kendon’s analysis is that communicative bodily movement has gestalt properties. For various reasons, a spatio-temporally extended stretch of movement suddenly stands out as a unitary expressive whole – a *stroke* – that is unified by some kind of meaning. This is so even in cases where a gesture is realized through a relatively complex set of movements. In actual situations, these gestural acts are recognized as they unfold in time, rather than as static units that are immediately given as wholes. The meaning and relevance of a gesture is different when it is ongoing, or held, than when there has been a retraction (Andrén 2011). Consequently, gestural gestalts, or action gestalts more generally, should be understood as gestalts in a *dynamic* sense.

Gestures and multimodal utterances may be further characterized in terms of the distinction between *weak* and *strong* gestalts (Köhler 1938 [1920]). Strong gestalts are holistic entities that, strictly speaking, do not have discernable parts at all. Weak gestalts result from a dialectic between parts and wholes, such that the parts make the overall entity what it is, while the overall entity also contributes to making the parts what they are. In the case of weak gestalts, the parts are perceivable as unitary phenomena in themselves, even though they are part of a larger whole. Strong and weak gestalts exist along a spectrum. Gestures are generally to be found towards the weaker end of the spectrum: people are not completely blind to their parts – i.e. the movements that make up gestures – even though they tend to see through them to understand the overall acts as meaningful wholes; in particular when it comes to the stroke phase of movement excursions. Multimodal utterances, consisting in both bodily movement and speech, are even more clearly a kind of weak gestalts. Note though that Köhler discussed ‘physical Gestalten’, which have gestalt qualities in an objective sense. Since I am dealing here with (inter)subjective and meaningful phenomena such as gesture, I should also say that gestures are *dependent* gestalts: they depend on subjective articulation to receive their gestalt-like properties. Gestalt psychology has traditionally been concerned mainly with perception, but it should be noted that the gestalt character of gestures, and other kinds of actions, is also constitutive of the organization of their performance.14

![Figure 1: Operation-act-activity interdependence.](image)

---

14 Koffka (1938 [1915]: 377) argues that action is not only gestalt-like in perception but also in the constitution of its *production*. He writes that ‘we may in fact place the experiencing of Gestalten squarely beside that of creating Gestalten; to sing or play a melody, dash off a sketch, write, and so forth, are not cases where one sings or plays tones, or draws or writes strokes. *The motor act is an organized whole process.*’
Leontiev’s (1981) tripartite distinction between operations, actions, and activity is useful for describing gestures and multimodal utterances in terms of a dialectic between parts and wholes. I will use these terms here, although in a somewhat unorthodox way that is not completely true to the way Leontiev originally defined them. Through some bending of their meaning, they can be made to fit well with the idea of weak dependent gestalts. Accordingly, one may distinguish a level of actions understood as the level at which people tend to perceive action both in the instrumental and communicative sense (see Figure 1). Gestures and other actions are united by a sense of purpose or meaning.

Below the level of actions lies a level that Leontiev calls operations. In the case of gesture, operations would be the movements that realize the act: its parts. In the case of single acts consisting of several discernable operations, each operation on its own would generally be perceived as an incomplete, perhaps incomprehensible, movement – not as an action. There is also a level above the level of action, termed activities. Activities consist of a number of actions with a more or less clear orientation in common: an overall motive. As I have already said, I am using Leontiev’s operation-action-activity distinction in an unorthodox way here. While Leontiev emphasized a principled dissociation between the levels, I prefer to use these terms in a way that does not downplay the potential dependencies between the levels (or the problems associated with forcing all sorts of complex situations into this relatively simple scheme).

Linell’s (1998, 2009) concept of act-activity interdependence addresses this partial dependence between action and activity. Their interdependence means that an activity partly becomes what it is through its constituent acts (its parts), and the constituent acts partly become what they are through their embeddedness in an overarching activity: a bi-directional dialectic between parts and wholes. For example, the utterance involved in pleading guilty in court is part of what makes the whole a trial in the first place. At the same time, the act of pleading guilty becomes valid because it is part of a trial. Drawing on Leontiev’s distinction, I propose that an analogous notion of act-operation interdependence be employed, or, better yet, that Leontiev’s threefold distinction be transformed into a more full-blown concept of operation-act-activity interdependence. It is clearly undesirable to treat any of the relations between the levels as unidirectional arrows, since doing so implies an infeasible form of reduction: either downwards, to underlying constituents; or upwards, to rigid forms of structuralism or an overemphasis on the role of context (as opposed to form) in the determination of an act’s meaning.

The concept of interdependence captures the way that individuals are active (agency), while at the same time acting in a way that is both constrained and supported by the more-or-less typified contingencies and conditions of (social) situations and activities. This is a view of agency that acknowledges both the initiating/creative and responsive/sensitive/responsible aspects of action. The

---

15 Many mainstream pragmatic theories, including those of Searle, Grice, and Sperber and Wilson, tend to ignore the holistic aspects of (communicative) actions as they try to derive their meanings exclusively from constituent acts (Linell 2009: 187).
overall position that emerges from the operation-act-activity scheme of human conduct can be characterized as one of *moderate holism* (cf. Linell 2009: 18). In the next section, all three levels play a role; but the main analytic focus lies on the interplay between the operation and action levels.

**EMPIRICAL ANALYSIS: ACTION COMPLETION IN SOCIAL SETTINGS**

The examples analyzed here are selected because they are all from situations where (a) some action has just been completed, and (b) there is some sort of *manifest expression of the status of the act as being completed*. These manifestations consist in both speech and bodily movement. All speech is translated from the Swedish except for the word *så*, which is of special relevance to the analysis and so explained as part of that analysis. The examples are presented as comic strips, which most readers will be able to read and understand more or less immediately. This form of representation works best where detailed transcription of speech is not so crucial. The inspiration for using comic strips to present analysis of interactional sequences comes from work by Oskar Lindwall and Jonas Ivarsson. (A methodological discussion on their use is given in Lindwall 2008.)

Figure 2 shows Tess (aged 26 months) playing with a comb and a doll, together with her mother. In the first frame, she performs a combing movement (a stroke) from a first-person perspective (Zlatev & Andrén 2009; cf. ‘character viewpoint’: McNeill 1992), in temporal and semantic coordination with the content-loaded verb ‘(to) comb’. The Swedish word form makes it clear that it is in infinitive form and the ambiguity between noun and verb forms of the English word ‘comb’ is an artifact of the translation. Apparently, Tess is familiar not only with the culturally established and typified act of combing itself, but also with the conventional word denoting the act. The explicit naming of the act through speech plays an important part in making the multimodal act, as a whole, stand out as a communicative figure: the spoken word highlights the specific moment in the movement excursion where something especially significant happens. The multimodal act is thus clearly brought into the *main attentional track* (cf. Kendon 1978, Goffman 1974) as an intersubjectively foregrounded entity. The performed act is *content-loaded* in the same way as children’s empty-handed iconic gestures. What that means is that the performer must have access to culturally standardized knowledge to be able to perform the gesture properly: how combs are used, what they look like, and so on. The act differs from a purely instrumental act because it is performed in a markedly reduced way: with a single, quick stroke the hair of the doll becomes ‘combed’. While the act performed is not realistic, it is still sufficiently similar to the type of act it invokes to be recognizable as

---

16 In several studies of children’s gestures, content-loaded gestures (which include iconic and conventional gestures) are distinguished from deictic gestures, where the latter can be used to refer to things in the world even when one does not yet have any clear conceptual knowledge of the thing being referred to or lacks the appropriate word for the thing being pointed to (e.g. Pizzuto et al. 2005). Deictic gestures are not necessarily ‘non-conceptual’, however (cf. Petitto 1987, Kendon & Versante 2003, Wilkins 2003). The point is that, in contrast to content-loaded gestures, they can be, to a larger extent.
a token of that type (cf. Bates et al. 1979: 39). Indeed, what seems to be at issue in this situation is precisely invoking the situation-transcending, standardized, and pre-established type. The reduced and unrealistic form of the performance may actually help to achieve that effect because of how it contrasts with the instrumental version of the act.

Immediately after performing the act, Tess puts the comb away on the table (Frame 2). In contrast to the act in Frame 1, putting the comb away is closer to the level of operations than to the level of action, because it mainly appears to be part of phasing out another action: the act performed in Frame 1. It is thus subsumed into an overarching action gestalt, being part of what constitutes the previous act as finished rather than being an independent act on its own. Four out of four persons shown the video clip described the sequence in terms of combing the doll and no-one mentioned that something was put on the table: evidence in favor of interpreting the putting away as a less salient or relevant aspect of the scene – one that tends to fall outside the main attentive track. (See the discussion about people’s consistency in judging the communicative status of movements in Kendon 1978.) This sort of putting away is different from putting something at a particular place while saying ‘there’, which would make the putting appear foregrounded as an action in its own right (cf. Clark 2003 on ‘place for’). It is crucial for the appearances of this act that the child is not (visibly) paying particular attention to where or how she puts the comb away.

At the same time, it would be wrong to say that this is just a case of a ‘mere putting away’ operation or ‘mere retraction’. The moment is highlighted through its temporal and semantic coordination with the Swedish word så. Thus, the word så takes part in the temporal (or aspectual) profiling of the act of combing, marking the shift from its ongoing performance to its status as ‘accomplished’. In the contexts analyzed here, the word så roughly means ‘like this’ or ‘like that’, similar in many respects to certain uses of the German so, the English like (this/that)17 (Streeck 2002) and so, and the Estonian nii (Keevallik 2005; 17 ‘Like’ is mostly used together with other words and cannot stand alone as a one-word utterance in the same way that the Swedish så can. In some cases ‘like’ is the only word in a clause; but then it is generally combined with a

Figure 2: Combing the hair of the doll (aged 26 months).
More generally, så is a highly multifunctional word that appears as a deictic word, discourse marker, adverb, conjunction, and more (Lindström & Londen 2008, Norrby 2002, Ottesjö & Lindström 2006, Ekerot 1988). There are a large number of variant forms residing somewhere between collocations and compounds: e.g., sådär (‘so-there’, often referring backwards in time), såhär (‘so-here’, often referring to an upcoming action), sådärja (‘so-there-yeah’), så att (s’att, ‘so-that’), också (‘and-so’, also), sån(t) (such), etc. In relation to gesture, these words have mainly been described in contexts of demonstration or ‘body quotation’ (Streeck 2002). In such cases, they are generally coordinated with the performance (the stroke) of a gestural action. However, the Swedish så is often used after the act: i.e., when the act is accomplished, rather than during its ongoing performance. In such cases, it indicates that something was properly accomplished: according to normative conventions or, sometimes, in relation to locally established expectations.

In Frame 2, Tess uses the word så in the sense of accomplishment, pronouncing it after the act. The word thus evaluates and concludes the act (shown in Frame 1) completed just a moment before; at the same time, it is uttered in temporal coordination with Frame 2’s operational move (putting the object aside after it has been used), which constitutes the end of the previous action of combing. Tess uses it in relation to non-locally established knowledge of the typified usage of combs: this is obviously not the first time she has encountered a comb and its uses. That this word is often used at the moment in time where acts are completed means it is often coordinated with retractions. This is different from the commonly adopted (McNeill 2005: 32) but not undisputed (Kendon 1985, 2004) rule of thumb according to which the nuclei of the tone unit in speech is typically coordinated with strokes. This occurs precisely in relation to the semantic theme of completion itself, which is evidence in favor of a Kendonian, more flexible, interpretation of the relationship between gesture and speech. The children studied here all clearly distinguish between the use of så in coordination with performance (stroke) and accomplishment (retraction), so that it is typically very clear which of the two is employed on a given occasion.

In Frame 3, the ‘real’ retraction comes: the one that ends in the rest position. Hence, the retraction phase following the stroke (Frame 1) turns into a biphasic (two-step) procedure (Frame 2 + Frame 3), which is different from the single-phase retractions of empty-handed gesture. This often happens for communicative actions that involve objects, where an ‘extra’ phase is required to disengage from the object. Precisely in these circumstances, the act of putting may appear as ‘mere putting away’ rather than as a foregrounded action/stroke in its own right. This shows how the vocabulary of gesture phases is useful for description of communicative acts that involve objects, even though elaboration is needed to make it fit with the particular conditions of such acts.

subsequent gesture, such as ‘an’ I was completely baffled [pause] like <pantomime of being baffled>, thereby adds a further qualification (cf. Slama-Cazacu 1976 on mixed syntax).
To account for the movement in Figure 2, it is necessary to introduce the notion of a disengagement phase. This movement has a logic distinct from retractions that end up in a rest position. Instead, another phase comes afterwards: e.g., a new preparation phase, a stroke, or, as in this case, a retraction phase. This is somewhat similar to Kendon’s (1980: 213) notion of a ‘partial recovery’ (partial retraction) in empty-handed gesture, where the hand stops at an intermediate home position from which there may either be (a) further gestural excursions before the hand finally returns to the original position of rest or (b) a retraction back to the original position of rest after a brief hold at the intermediate position. However, unlike retraction phases (partial or otherwise), disengagement phases need not end up in a position of rest (they rarely do). At the end of a disengagement, the arm and hand may be fully extended, which is not characteristic of any sort of position of rest or home position of the kind that goes together with retraction/recovery. Neither need there be any stopped movement at the end of the disengagement phase, as the hand may keep moving after the object is put down. Disengagement is further different from a partial recovery because the disengagement is object related – enforced by the handling of an object – whereas a (partial) recovery may come about for rhetorical purposes, such as closing intermediate units of activity in a hierarchically organized communicative activity. At the same time, disengagements are similar to retractions. They often appear as part of the disattend track: not as a focal action (stroke) in itself but rather as a kind of phasing out of the action.

Figure 3: Betty (aged 24 months) is ‘serving coffee’.

Figure 3 shows a similar situation. Betty first performs a content-loaded, culturally typified action (the ‘stroke’, Frame 1) with the canonical type of object (albeit a toy version of it) from a first-person perspective, in a gesturally reduced way: the whole movement is snappy, without a pause for pouring. Again, it is performed in temporal and semantic coordination with a content-loaded word (in this case, a noun rather than a verb) that relates to the typified identity of the act. Frame 2 shows another example of a

\footnote{Kita \textit{et al.} (1998) make use of the highly similar concept of partial retraction, which they define as a retraction interrupted for another preparation phase or another stroke. That is, they require that another gesture stroke comes afterward. This makes their partial retraction more different from disengagement than Kendon’s partial recovery. Also note that there is no ‘interrupted’ quality built into the concept of a disengagement phase.}
disengagement phase, this time without the word så. However, before the final retraction to home position (Frame 4), another intermediate retracting movement (Frame 3) is performed: characterized by effort and shape and coordinated with the word så. (‘Effort and shape’ is what defines a stroke: see Kendon 1980). The effort and shape in this case consist in a quick backwards whipping movement, with the wrist tilting backwards. The retraction in Frame 3 goes slightly upwards towards the shoulder, rather than towards a position of rest. The movement ends in a brief hold, before finally proceeding to the rest position shown in Frame 4. The retraction in Frame 3 is a hybrid between a (partial) recovery (it is a movement away from the stroke articulation) and a stroke phase (it has properties of effort and shape), justifying the introduction of another term: marked retraction. Again, the word så appears in temporal and semantic coordination with an operational movement, constituting the marked retraction, that manifests the completedness and accomplishment of an action. As in the previous example, the part of the excursion that goes together with så has the status of an operation: not an action. As such, it appears mostly as a subsidiary operational part of the main act’s gestalt (Frame 1), although it is marked with (some) effort and shape. Once again, the result of combining a marked ‘extra’ retraction with the evaluative and temporally backwards-pointing så is an unfolding profiling of the aspectual character of the main act, marking the gestalt shift from ‘ongoing performance’ to ‘accomplished’. By Frame 3, the communicative act in Frame 1 has become accomplished, rather than being an ongoing performance. Still, even if the retraction in Frame 3 is less explicit and less action-like than the act in Frame 1, it is more explicit than the ‘mere’ retraction in Frame 4. To conclude: in this case, the retraction is expanded into a three-step procedure: disengagement (Frame 2), marked retraction (Frame 3), and retraction (Frame 4). In the previous example, the manifest character of the completedness was mainly achieved through the word så. In contrast, in this case the completedness of the action is also made manifest through the peculiar (marked) retraction in Frame 3.

In the next example, Howard has just performed a whole sequence of preparatory acts, subsumed under the overarching motive of making the train ready to go, before the performance shown in Figure 4. These
preparations involved locating a toy figure, putting it on top of the front car of the train (as the train driver), and rotating the head of the figure to make sure that it faces the direction of the train. After these preparations, the train is finally ready to depart. In Frame 1, Howard grabs the toy figure and moves the train briefly forwards. This movement is performed in temporal and semantic coordination with the content-loaded verb ‘drive’. The act as a whole is obviously dependent on culturally defined and typified knowledge of what trains do: that they have drivers, that the drivers should sit in the front of the train, that the drivers need to look in the direction they are driving, and so on. (To what extent this typified knowledge is knowledge about how to use the toy train, or knowledge about ‘real’ trains, is a different issue, with no immediate bearing on the arguments put forward here – it is cultural knowledge in both cases.) In this example, too, the main act is performed in a very reduced way: Howard seems to consider the act complete as soon as the train has moved a little and the word ‘drive’ has been pronounced. Like the previous examples, what matters seems to be to invoke the idea of a certain type, rather than providing a realistic performance of the type.

What is different here, compared to previous examples, is that when the act is performed, the object involved is already supported by the table. No ‘extra’ disengagement phase is therefore necessary. Howard does not, however, embark on a retraction to a position of rest. His hand does retract: it moves back towards the body after completing its job of driving the train; but, rather than ending in a position of rest, it ends in an audible hand clap (Frame 2). Just like the marked retraction in Figure 3, it is more an operation than a referential action, although it is a relatively foregrounded type of operation. Once again, the stroke-like properties of the marked retraction has the character of profiling the act just performed. It marks the shift from ongoing performance to accomplished performance. By virtue of being more an operation, it primarily makes sense relative to the main act it concludes: i.e., as one of several parts of an overall action gestalt. Hand claps as parts of utterances often appear in contexts where something is completed (and another thing may follow).

![Figure 5: Completedness and audible gestures.](image-url)
Adults make use of audible elements of gestures in functionally similar ways. Figure 5 shows an adult who performs a gesture on the semantic theme of **DONE**, in Frame 1. As part of the retraction, in Frame 2, she produces a clapping sound much as Howard did.

Note that Howard’s hand clap is coordinated temporally and semantically with the word **så** (in the accomplishment sense, as in the previous examples). Afterwards, a ‘real’ retraction comes: his hands retract to a rest position (not shown). It is unclear whether Howard’s **så** refers to accomplishment of the driving in particular, or of the whole sequence of acts (including the preparatory ones) of which the driving was the culmination. Perhaps it is most correct to say that this token of **så** is undifferentiated in this regard.

**Figure 6: Amy makes the bed (aged 30 months).**

---

19 I am grateful to Simon Harrison for providing this example. See Harrison (2009: 125-132) for an extensive analysis of the same example, albeit with different questions in mind.

20 I have observed several university teachers using hand claps in the context of shifting from ‘before lesson has started’ to ‘the lesson has now started’. In several of these occasions, the claps were coordinated temporally with the word **så**, sometimes as part of a larger linguistic construction such as **‘så+[CLAP], let’s start’**. As with Figures 4 and 5, the claps mark the transition between segments of the activity.
Figure 6 provides evidence for distinguishing the use of *så* for the accomplishment of a single act and the accomplishment of a whole sequence of acts. Frames 1–3 show Amy performing a series of actions, all oriented towards the overall motive of making the bed. Each act results in covering some part of the bed with the bedspread. For each of the three action strokes, she says *så* – but in the performance sense rather than the accomplishment sense. As mentioned before, when words like *så* are used in the performance sense, the temporal coordination with the stroke of the action is part of what makes the meaning of the word recognizable, in a deictic manner. The result of the action in Frame 3 is that the bed is now fully covered by the bedspread. At that point she takes a step backwards (Frame 4) while saying *så* in the accomplishment sense. Here, the word *så* seems to be directed at the accomplishment of the whole sequence of acts, not just the most recent one. Remember that the acts in Frames 1–3 already had a *så* associated with each one of them. While saying *så* in a loud voice (Frame 4), Amy moves her whole body backwards in a joyful dance-like step, which simultaneously indicates her satisfaction and the accomplishment of the motive of the overall activity that she was engaged in. In this way, she demonstrably orients herself to the hierarchical organization of the bed-making activity. Once again, her overall motive is a culturally defined one, resting on conventionalized knowledge of what a bed should look like when it is properly covered with a bedspread, and how this is accomplished. To be able to intend to make the bed, and to be able to express this manifestly, she must have access to this particular knowledge.

In Frame 5 she inspects her deed. While she is doing that, an adult in the background asks if she is ‘done?’ with the bed, thereby also displaying an orientation to the performed sequence of actions. When Amy replies in Frame 6, she takes yet another step backwards and quickly folds her arms behind her back, saying ‘yes’ while nodding affirmatively, all at once. In this way, she places herself in a position, or *stance*, with markedly low potential for further action, in contrast to the more neutral posture she held just a moment before (Frame 5). Clark’s (2003: 261) ‘accessibility principle’ is relevant here, although in an inverse way:

*Accessibility principle*: All other things being equal, an object is in better place for the next step of action in a joint activity when it is more accessible for the vision, audition, touch, or manipulation required in the next step.

By moving backwards, Amy makes the bed less accessible. She therefore appears more ‘done’ than ready for a next step.\(^{21}\) The folding of the arms behind the back has the stroke-like properties of effort and shape,

\(^{21}\) Moving backwards in gesture seems to be something like a general semantic theme of passivity. Gestures such as *GIVE_UP* (both where the hands are held high in the air and where the hands are placed on top of the head or behind the neck) derive their meaning from the reduced action potential of this position. Kendon (personal communication; see also Kendon 2004: 251–255) observes that, when people hold up a vertical flat hand to indicate ‘stop’, they usually do it with a movement towards the *other* if the stopping concerns the *other*; but a movement towards the self if the stopping concerns themselves. Green (1968) lists a number of conventional gestures, with meanings such as ‘disclaimer’, ‘apology’, ‘cancellation’, and ‘lack of concern’, all of which seem
in contrast to the more relaxed position she held just a moment before (Frame 5). Although this move acknowledges the status of the bed as having been made (i.e., covered by the bedspread), it seems counterintuitive to describe it as a referential gesture on the action level. It is probably better understood on the level of operations, in part because this move seems mainly to function as a concluding part of Amy’s previous action gestalt. This move is a constitutive part of, and makes intersubjectively manifest, the progressively emergent structure of the activity and it makes her previous sequence of actions appear in a certain way: namely, as accomplished, finished, done.

DISCUSSION

The above analysis shows how, in many cases, the proper completion of various acts is not just publicly available (as something that could be seen by someone who happens to be attending to it) but actively made intersubjectively manifest. The children achieve this, in part, by moving their bodies in certain ways. I argued that, in the specific context of ‘being done (with something)’, retraction phases sometimes take on the stroke-like properties of effort and shape. They thus constitute a hybrid case between strokes and retractions, which I have called marked retractions (a term coined for this paper). In such cases, the retraction phase following a foregrounded and communicative action is highlighted by e.g. performing it quickly, ending it with an audible hand clap, folding one’s arms behind one’s back, or taking a step backwards. Retraction to a position of rest does not appear until after the marked retraction, as a separate movement phase.

The analysis mainly concerned communicative acts that involve the handling of objects. One may note that marked retractions may appear as part of empty-handed gestures too, as shown in Figure 5; so it is not only a matter of gestural actions that involve handling of objects. The focus on communicative acts that involve objects led to the introduction of the notion of disengagement phase, as an elaboration and extension of Kendon’s (1980, 2004) vocabulary for describing empty-handed gesture phases. The disengagement phase consists in the ‘mere’ disengagement from an object (one that happens to be held in the hand), without this disengagement being foregrounded into the main attentive track as are stroke-like movements. Unlike the retraction phase, the disengagement phase does not end in a position of rest. The arm may be fully extended, and the hand may be far from the body after its performance. The hand need not make any temporary pauses (holds) in its movement when the object is put down – although it can. Typically, the manual movement only changes direction once the object is put down. Communicative acts that involve objects differ from each other with respect to whether or not there is still an object in the hand after the performance of the act. For this reason, they also differ with respect to whether or not they require a disengagement phase afterwards. The examples analyzed above reveal various constellations of marked partly to derive their meaning from a reduced action potential in some form – either through some form of withdrawal, or by a show of empty hands.
retractions, disengagement phases, and ‘proper’ retractions. In one example (Figure 3), all three are present as separate movement phases.22

In all of the examples, speech also has a role in making action completion manifest. The Swedish word så is involved in all of them, except for the example involving only adults. Så has a range of meanings. It is used here mainly to mean ‘like-that’ – coming after a performed act, highlighting its accomplishment. In the final example, it is also pronounced during the performance of the act, to mean ‘like-this’ – highlighting the apex of the performance. Many other verbal means of commenting, evaluating, naming, etc. exist, sometimes serving similar purposes of making action completion manifest. Nevertheless, så remains one of the more common, and specialized, resources used for such purposes. It is a typified, socially shared, conventional resource. All four children analyzed here had learned to use it to make accomplishment of actions manifest, in a highly generalized way: i.e. not only with a particular type of action.

One peculiar feature of the coordination of this particular word with bodily movements is that it often comes in temporal coordination with some sort of retraction, rather than a stroke. This observation is in line with a Kendon-inspired understanding of the relationship between speech and gesture, which emphasizes the flexible and mutually adjusted nature of this relationship (see Kendon 1985, 2004) – adapted to the particular expressive needs at hand. Although the bodily manifestations of action completion that I described in the analysis lie closer to the level of operations, the word så is clearly a referential action in its own right. Kendon (2004: 7) notes that participants in social interaction always seem to treat speech as part of the main attentive track. Still, så means little when taken in isolation: it is highly dependent on being part of the overall action gestalt of the ‘main act’ it deictically refers back to – including its precise timing relative to the act (during versus afterwards).

I argued that the relation between the parts and wholes of gestural movements and whole multimodal ensembles is best understood in terms of a moderate holism, according to which gestures are weak gestalts located at the action level (although so-called beat gestures are often rather on the operation level). They possess the meaning they have by virtue of being located in the middle of an operation-act-activity interdependence nexus: an extension of Linell’s (1998, 2009) concept of act-activity interdependence, with inspiration from Leontiev (1981) regarding the introduction of the operation level.23

The analysis focused mainly, but not exclusively, on the interplay between the operation level and the action level. That said, the interplay between all three levels should be acknowledged. This stands in

---

22 Marked retractions can occur either with empty-handed retractions (either partial or not) or with disengagements. Therefore, they represent not a separate type of retraction, but rather a modulation of the other types – one that adds the properties of effort and shape to the performance.

23 Kendon (1972: 191, with reference to Scheflen 1964) notes that shifts in posture commonly signal shifts between ‘major units of communicational activity’: i.e., the operations level (posture shifts) and the activity level may be directly interdependent.
principled contrast to any reductionist derivation of gesture/action meaning solely from constituent form features on the operation level – or solely from the relevancies of the overall social activity of which the acts are part. It also stands in contrast to thinking of gestures as ‘closed units’: i.e., atoms of meaning whose unfolding performance is already determined by the performer at the moment of embarking on a preparation phase. The possibility exists of adding – on the operational level – a further move to the overall action gestalt during the course of the performance. This may affect the perception of the action as a whole. Further, as I have shown elsewhere, gesture performances can be sensitive to the way others respond to the gestures while they are being performed (Andrén 2011): a clear demonstration of their partly open nature.

All the manifestations of action completion analyzed here – i.e., marked retractions of various sorts and use of the word så – have the effect of making the aspectual profile of the ongoing action intersubjectively manifest: namely, they mark the shift from ongoing performance to accomplishment, and they do this at the precise point in time when the shifts occur. To some extent, they could even be said to constitute these shifts as accomplished social facts. Even though they are (in some sense) performed separately from the main acts, they still contribute to making those main acts what they are.

Given the power of the word så to make accomplishments manifest, it is interesting to note that parents frequently use så – in the accomplishment sense – to evaluate their children’s acts, thereby making the normatively proper and typified character of the acts manifest for the children.24 Even though the affordances of objects in children’s life-worlds are generally constructed by humans with usability in mind (Sinha 2009: 294), many artifacts and other objects still ‘do not tell us what they afford’ (Zukow-Goldring 1997: 213). Someone else must provide guidance.25 Parents may use så precisely to help their children overcome the failure of the objects’ affordances to provide the knowledge required to participate in, understand, and attend to the social world in which these objects are used (cf. assisted imitation, Zukow-Goldring & Arbib 2007). One of the main reasons why the other, and interactive engagement with the other, must be an explicit and developed part of a comprehensive theory of the genesis of meaning is the non-obvious character of much of the cultural knowledge people operate by – even take for granted. (Recall the pure action-perception accounts of thought and action critiqued in the beginning of this article.26)

24 In the several hours of data studied here, such parental uses of så occur, on average, once every fourth minute. Of course, this is only one of many ways to give feedback to children regarding what they do. Note also that parents use this word even more frequently to mark up their own actions.
25 This is a variant on Vygotsky’s themes of scaffolding, zone of proximal development (Vygotsky 1978), and education of attention (Vygotsky 1981).
26 Following Sartre, one might say that language – and, more generally, interaction with the other – teaches one the structure that one’s body and one’s actions have for the other (cf. Schutz 1948: 192). This is an indirect form of embodiment, often forgotten in accounts of embodiment’s role in the genesis of meaning (e.g., Johnson 1987).
All the children’s expressive acts analyzed in this paper are clearly oriented towards socially typified knowledge. Here, Schutz’s theorizing on typification becomes particularly useful. In the analyzed examples, typified procedures such as using a comb, serving coffee, driving a train, and making the bed all manifest themselves as the children’s intended meanings. Clearly, these intentions are not the original creations of the children, but rather typifications, adapted to the situation at hand in an improvisational manner. In the several hours of data on which this study is based, there is no instance where a child says så after performing an act that would, according to cultural standards, count as a failed or faulty action. Observers can understand these actions, because they can ‘re-cognize’ the meanings: that is, the actions invoke meanings that people are already familiar with at some level. (Recall Schutz’s discussion of Verstehen.)

Typification, or conventionality – on the side of the signified (substance, or content), rather than on the side of the signifier (form, or expression) – is a neglected issue in gesture research, presumably because the domain of meaning is typically conceived as completely separate from the ability to gesture, despite the logical dependence of content-loaded gestures on such knowledge. This dependence on typification, on the side of the signified, is particularly salient in iconic action-based gestures: bodily actions (gestures) that refer to bodily actions. Such gestures happen to represent the overwhelmingly dominant type of iconic gesture in young children (Andrén 2010). When a culturally typified action is communicatively re-enacted as an action-based gesture, the signer side is partly conventional too: the typified properties of the culturally established action bleed over to the signifier side. This is so because the modality of the signifier is the same as the modality of the signified and hence draws on the same knowledge. As noted earlier, this need not mean that the enactment of a culturally established action type is of a ‘realistic’ kind, but signifier and signified are similar enough for the signifier to be recognizable as a token of this type – at least with some support from the contextual framing and co-occurring speech. Co-verbal gestures are commonly claimed to be, for the most part, spontaneous and idiosyncratic (McNeill 2005: 91), not conventional: a somewhat homogenized conceptualization of their nature. The claim may be true to some extent as a characterization of gesture in contrast to full-fledged (spoken or signed) language. Still, it is striking how little of the content of gestures is invented on the spot in children. The relevant literature shows that children have ‘much difficulty understanding... creative iconic gestures’ (Tomasello 2008: 149, italics mine). Surprisingly often, it is possible to track children’s iconic gestures to imitations of adults’ typified actions and gestures.

Mirror neurons are sometimes invoked to explain the transparency of interpersonal action and gesture understanding. Although mirror neurons seem to be an important piece of the general enabling capacity for such understanding, it would be a mistake to assume that mirror neurons – by themselves – provide the substance of the understanding (cf. Andrén 2010: 214 for a more extended version of this argument). Arbib et al. (2005: 239) write, ‘mirror neurons are not innate but instead correspond to a repertoire of learned
actions and learned methods for recognizing those actions’ (see also Heyes 2010). Calvo-Merino (Calvo-Merino et al. 2005) has shown that such obviously culturally defined activities as ballet and capoeira dancing shape mirror neurons differently in practitioners of each of those two cultural practices.

The findings are consistent with a view of gesture and action as skilled performance (cf. Kendon 1990: 4) that makes reference to particular and diverse forms of knowledge (including previous experience) in their constitution. Echoing Schutz, the iconicity of content-loaded gestures that invoke interpretation as typified action seems to be strongly dependent on socially typified experiences of various sorts, both in their production and comprehension. I believe this to be a weaker type of conventionality than the rule-like conventionality (normativity) found in e.g. the core principles of grammar – but it is still a form of conventionality (cf. Andrén 2010: 54).

Even though many aspects of gesture are not as strongly conventionalized as various elements of sign language, this does not mean that co-verbal gesture should be characterized as non-conventional. (See Andrén 2010: 46; Harrison 2009; Kendon 1996, 2004, 2008; and Streeck 2009 for critical remarks on treatments of co-verbal gesture that overly downplay its conventional aspects.) I use the term ‘typification’ here in the sense of mutually shared and conventionalized knowledge. Typified conventions need not be strict enough to qualify as rule-like normative conventions, dictating obligatory aspects of the performances, deviations from which are simply wrong (like so-called Emblems: cf. McNeill 1992; e.g., the thumbs up gesture or the V-sign); but they also need not be weak enough to qualify as ‘mere’ regularities or vaguely ‘normal’ ways of acting. They are indeed known, in common, as identifiable types. Recall that the children studied here were able to use culturally established labels to name the acts they performed – or relevant aspects of those acts. Typified conventionality may be found, not just bluntly ‘in the gestures’, but, more precisely, (a) in the signifier, (b) in the signified, and (c) in the relation between signifier and signified. Other conventional aspects surrounding the use of gesture may be better understood as residing outside the relation between signifier and signified (cf. Kita 2009). Gesture researchers need to be more careful, and not provide dichotomous statements that one or another gesture or communicative act either is or is not conventional. (For further discussion of different kinds and levels of conventionality in relation to gesture, see Andrén 2010).

This article has highlighted how the state of being done with something may be made intersubjectively manifest through bodily movements and through the semantics and timing of speech, and how this is often strongly oriented towards culturally established conventions relating to the typified uses of various objects – without which many of these expressions would not exist.

Acknowledgements

I am grateful to the reviewers for this article, including Adam Kendon, for their comments; and to Jordan Zlatev, Per Linell, Simon Harrison, Oskar Lindwall, Paul Cibulka, and Petr Kaderka for their comments.
and discussions. I would also like to acknowledge Erik Philip Sørensens stiftelse för genetisk och humanistisk forskning for part of the funding, and the Centre for Cognitive Semiotics (CCS) for providing funding and a stimulating interdisciplinary environment for research on the relations between different semiotic resources.

REFERENCES


