Introduction

Multimodal (Inter)action Analysis (Norris, 2004, 2011; Pirini, 2014) is a methodological framework which developed out of an empirical necessity. It was driven by theoretical demands to account for additional modes of communication in the qualitative analysis of real-time interaction and by new video recording technologies and data collection practices in anthropological and applied linguistics. In the early 1990s, discourse analytic approaches began to adopt social semiotic perspectives alongside inventories of methodological tools for the analysis of multiple communicative modes in textual phenomena. However, empirical approaches using ethnomethodological and video-ethnographic data collection methods (see Wei, this volume) lacked an approach to inductively analyse other communicative modes which emerge in interaction like that of gaze, posture, proxemics and gesture.

It is not as if pioneering work had not been conducted on non-verbal modes of communication. Influential figures like Ray Birdwhistle, Albert Scheflen and Adam Kendon (along with many others working together at the University of Pennsylvania and the University of Chicago) had been generating new theoretical insights and developing methods of transcription for para- and non-linguistic communicative modes. However, given the substantial void in understandings about non-verbal communication, the vast majority of early work in kinesics sought to describe and explain the communicative contribution of individual modes (see Kendon, 1967; Scheflen, 1964) or bi-modal relationships like speech–gesture links (Kendon, 1980). The majority of the linguistics community at the time dismissed kinesics as not truly integral to the study of language, likely due to the lack of theorisation regarding relationships between multiple modes of communication. Multimodal (Inter)action Analysis provides some of the key methodological tools to address the challenges faced when analysing non-verbal and verbal modes of communication together.

Given that all modes have different materialities and structures, a singular unit of analysis was arguably the most challenging methodological hurdle to overcome. However, it was also paramount. The single unit of analysis needed to be applicable across the broad spectrum of communicative modes to facilitate inductive qualitative analysis without implicitly allocating value to any single mode a priori. The framework itself also had to be able to account and

The mediated action forms the theoretical backbone of the Multimodal (Inter)action Analysis framework and operates as an analytical starting point. Through the mediated action, the so-called individual/society antinomy is finessed (Scollon, 1998) and the individual, the sociocultural and the environmental are recognised as always and inextricably intertwined and mutually influencing. Human beings are thus conceptualised as inextricable from the physical world in which they live. The traditional ‘inside’ vs ‘outside’ or the ‘me, others, environment’ distinction is recognized as problematic since sociality, the physical world and the individual are mutually constitutive. Even the action of perception is bidirectionally influenced and a result of interaction with the environment. Thus, rudimentary forms of cognition are in essence interactive. This important theoretical point explains the (inter) in (inter)action. All action is taken with and through the environment and thus all forms of action are inherently interactive.

The utility of the concept of mediated action has been brought to the fore in Multimodal (Inter)action Analysis through a subtle but significant partitioning between lower-level actions and higher-level actions which facilitates the utilisation of a single unit of analysis across a wide diversity of communicative modes. The importance of cross-modal applicability cannot be overstated. This ensures that inductive analysis does not allocate communicative salience to any single mode a priori. In the analytical procedure, all modes are in a sense ‘equalized’. It is only through analysis that communicative salience or the contribution of any single mode can be determined. This important feature of the methodological framework distinguishes it from methods which take linguistic action as the analytical departure point. When language and not action is in the analytical foreground, non-verbal modes are implicitly ‘secondary’. This is an unsustainable ontological position as almost any rigorous analysis of social interaction will reveal.

Since the original development, empirical enterprises employing Multimodal (Inter)action Analysis have generated a wealth of insights both theoretically in the ongoing development of Multimodal Mediated Theory (Geenen, 2013a, 2014; Norris, 2013) and analytically with the development of various methodological tools. Multimodal (Inter)action Analysis as an integrated framework provides tools for analysing the attention/awareness of social actors in any specific site of engagement through modal density (Norris, 2004, 2011), the production of modal configurations (Norris, 2016, 2017; Pirini, 2016, 2017) which help analysts discern the salience of particular modes during any segment of interaction and methods for the analysis of actions as related across time and space to various scales of action (Norris, 2017). The concepts, methodological tools and protocols outlined herein should equip burgeoning analysts as well as experienced researchers with the basic foundations to begin micro-analysis of real-time audio-video data.

**Mediated action**

Multimodal (Inter)action Analysis is theoretically indebted to Wertsch’s (1991, 1998) Mediated Action Theory and Scollon’s (1998, 2001) Mediated Discourse Theory, both of which were historically influenced by Vygotskyan socio-historical psychological approaches (Vygotsky, 1978). Central in all three are the notions of action and mediation. As Wertsch (1991) explains, “[w]hen action is given analytic priority, human beings are viewed as coming into contact with, and creating, their surroundings as well as themselves through the actions in
which they engage” (p. 8). A priority on action coupled with the acknowledgment that all human action is mediated by either psychological and/or physical tools ensures the individual/environment relationship is maintained in the basic analytical unit.

For Vygotsky (1991, 1998), the mediated action constitutes the most useful ecological unit of analysis on the basis that the unit itself maintains as much individual, interpersonal, sociocultural, historical and institutional complexity as possible. In any single action and the psychological and/or physical tools which mediate that action, various permeations of the individual, their sociocultural development and cultural situatedness emerge. The centrality of mediational means of varying origins, materialities and structures is paramount as the tools (cognitive, psychological, physical, etc.) invariably shape the nature of any action in consequential ways. The nature of any single action is not reducible to the individual or the environment in isolation but a result of the irreducible tension between the two. Furthermore, in many cases the particular nature of any action is a result of the mediational interrelationship (Geenen, 2014) as many actions are mediated by multiple mediational means simultaneously involving physical objects in the world, the physical body and cognitive cultural tools like language or mathematics.

While Vygotsky’s (1991, 1998) treatment of the theoretical value and analytical importance of the mediated action involved some discursive orientations, Scollon (1998) adopted the notions applying them directly to spoken language-in-interaction in the development of Mediated Discourse Theory (Scollon, 2001). Scollon’s argument regarding the utility of the mediated action as a unit of analysis echoes the anti-reductionist ethos of Vygotsky, and he extended this directly to interactional sociolinguistics on the grounds that discourse is best conceived of as a form of social action. Scollon’s assertion regarding discourse and adoption of the mediated action as a theoretical notion and analytical unit at first seems reminiscent of the same basic proposition of speech act theory insofar as the utterance is recognized as an action. However, as stipulated earlier in this chapter, it is primarily in the irreducibility of the social actor and mediational means where we find such comprehensive utility in the domain of linguistic action.

For example, an utterance can be considered a material mediated action. When using the mediated action as an analytical point of departure, the utterance is conceptualised as being influenced and affected by multiple simultaneous mediational means at a unique site of engagement.

The social actor’s speech organs or articulators produce the quality of the utterance. Their individual lexical decisions and idiolectical dispositions produce the lexical string and prosodic contours of the utterance. The utterance is also situated within a dialogue or exchange. There is some interactional or discursive exigency, which is undoubtedly influenced not only by the interaction itself and trajectory of the discourse but also simultaneously by interpersonal relationships and identities being produced, managed and altered at that specific moment in time.

Furthermore, the shaping of the social actor’s lexicon and vocabulary though seemingly individual and idiosyncratic is undoubtedly sociocultural. Any single word has been heard, learned and acquired through interaction with other social actors and other cultural tools such as books and signs. The words that social actors use are contingent on semantic enrichment through interaction, frequency of exposure and personal utility in juxtaposition with the other potential words or phrasings which could have emerged.

The mediated action in the form of a verbal utterance is additionally influenced by institutional and historical trajectories. The speech genre influencing and being co-produced at that time – as well as the street, office, café or house and their physical and cultural qualities – have bearing on the material nature of that particular action. One may be forced to yell over the buzz of passing traffic on the street or whisper so as not to disturb others in the library. That seemingly mundane utterance and the mediational means and cultural tools through which it
is undertaken is permeated by complex and interrelated individual, interpersonal, sociocultural and historical trajectories. This is distinctly how the mediated action maintains in one single unit, as much complexity of the whole as is possible.

Action as an analytical starting point has distinct theoretical advantages when employed as an analytical unit for a single mode of communication. Gestures can be conceptualised and analysed as mediated action, spoken discourse too as action. However, the key utility of Multimodal (Inter)action Analysis is its application to solve a difficult analytical impasse, which is borne by the fact that human beings never act through a single mode of communication. Through a subtle partitioning of the mediated action, Multimodal (Inter)action Analysis facilitates analysis of all modes of communication, without attributing analytical primacy to spoken language prior to analysis.

**Modes, lower and higher-level actions**

The challenge facing multimodal analysis comes from the varying structures, organisations and material longevities of modes of communication. For example, spoken language is combinatorial and auditory, and it has fleeting material longevity. On the structural side, natural languages allow for the combining of morphemes to create words and the combining of words to create sentences. Often, when approaching the notion of ‘meaning’ on a semantic level, the meaning of a string of words is fundamentally dependant on the semantic scope of the individual lexical items themselves. One can think of each as contributing some portion to the overall meaning of the whole utterance. In contrast, gesture is what McNeill (1992) describes as global-synthetic. *Global* references how determining the meaning of a gesture does not occur in a bottom-up fashion in the same way that individual lexemes combine to produce an utterance. To determine the meaning of a gesture, one does not consider position in gesture space, shape of individual fingers, path, speed and trajectory of stroke combining to result in the meaning of that gesture. The meaning of the individual features of a gesture are determined by the meaning of the whole and not the other way around. *Synthetic* refers to how various semantic properties may be couched in one representative form, in contrast to the *combinatorial* nature of language. Similar variations can be found across posture, gaze and proxemics. Furthermore, the basic unit of analysis in any single mode is not amenable to application in another. We find no stroke (gesture) in gaze and no utterance (spoken language) in posture. Norris’s (2004) partitioning of the mediated action artfully overcomes this dilemma.

The notion of mediated action is theoretically sound, but methodologically it is best to conceive of lower-level mediated actions and higher-level mediated actions. A lower-level action is “the smallest pragmatic meaning unit of a communicative mode” (Norris, 2004, p. 8). Given the variable material and organisational structure of modes previously described, a lower-level action in the mode of gesture would be the stroke or stroke hold. A stroke “is the gesture phase with meaning” (McNeill, 2007, p. 33) and stroke holds “are strokes in the sense of meaning and effort but occur with motionless hands” (McNeill, 2007, p. 33). As obligatory units insofar – as without a stroke/stroke hold, there is no gesture – these are the smallest pragmatic meaning units in the mode of gesture. Similarly, for posture the smallest unit is the point which Scheflen (1967) originally described as the postural unit bracketed by subtle shifts in head–eye orientation. In each case, these smallest units of any individual mode are usefully conceptualised as lower-level mediated actions. In real-time social interaction, lower-level actions occur rapidly and simultaneously across a wide variety of communicative modes. Additionally, these chains of lower-level actions build various co-produced higher-level actions.
A higher-level action is built through chains of lower-level actions; this is a co-produced larger scale action with socioculturally constituted and recognizable beginnings and endings. A higher-level action may be a conversation on the street between friends initiated through mutual recognition exemplified through the modes of posture, gaze and head movement. The ending may be co-produced through a wave in departure and customary words of farewell. A higher-level action may also be a specific topic which emerges during said conversation between friends on the street. In this case, the higher-level action may be opened or initiated through a topical shift in the discursive trajectory through the mode of spoken language, but it is often also exemplified through postural alterations, gaze shifts and proxemic alterations. Thus, higher-level actions nest within one another at multiple levels, and across multiple time scales.

Central to both lower-level actions and higher-level actions is that these units are heuristic in nature and specifically for analysis. In real-time social interaction, human beings do not produce utterances through the mode of spoken language or produce multiple chains of lower-level actions in the co-production of a higher-level action. Human beings act, interact and do things during the ever-changing unfolding of waking life in real time. The units described herein are thus tools, useful for partitioning the complexly orchestrated multitude of actions undertaken during lived experience (see Geenen, 2013b, for detailed discussion).

The notion of mediated action has been fruitfully applied to the physical environment and the way social actors allocate salience in their engagement with and perception of physical entities. This focus on the actor environment led to a conceptualisation of physical objects not just as entities occupying the same physical space as the social actor, but as constituting frozen actions. Objects in a physical surround can be conceived of and are often perceived as past mediated actions frozen in time, hence ‘frozen action’. A magazine lying open on a coffee table alongside a warm cup of tea represent the physical remnants of another social actor’s (inter)action with those objects at some other time–place. The objects themselves and their physical arrangement are indeed perceived and interpreted through mediated action; anybody entering that room is immediately aware of the action of leisurely reading a magazine with a warm cup of tea. Thus, objects themselves and the environment speak to the actions and interactions which produced their precise materiality. This notion of frozen actions provides the analyst with a unique method to analyse physical entities at a particular site of engagement with the same focus on mediated action.

Modal configurations

As with lower-level actions, higher-level actions can be analysed for the meaning they produce. Using the notion of modal configurations (Norris, 2009), the relative contribution of lower-level actions to the meaning produced through the higher-level action can be determined. Modal configurations refer to the hierarchical configuration of modes within a particular higher-level action. The analysis of modal configurations relies on first assessing the primary meaning of a higher-level action, and then determining which lower-level actions are hierarchically most important to the meaning the higher-level action produces. Using this approach, Norris and Pirini (2017) demonstrate how agreement and disagreement operate multimodally in a knowledge communication task. Here, disagreement was co-ordinated multimodally, but disagreement expressed through language lagged behind disagreement expressed through other modes. In other words, the communicative salience of particular modes unfolding at that site of engagement revealed that lower-level actions through non-verbal modes like posture, head movement and gaze were hierarchically more salient in the co-construction of disagreement, prior to the expression of such through verbal language.
Modal density: analysing attention and awareness in interaction

The related notions of attention and awareness capture the way that people shine the light of attention directly on some actions, while diffusely illuminating others. This light produces a focus of attention, which reduces as actions move away from the focus. In Multimodal (Inter)action Analysis we use the notion of a continuum of interactional attention/awareness (Norris, 2004) to explore how actions can be positioned closer or further from the focus of attention, and how actions can shift between different levels of focus. For example, when we see two people lining up for coffee, their actions demonstrate a focus on each other, with ordering coffee somewhere in the midground of their attention/awareness. Then, as the line approaches the counter, the focus of their attention shifts towards placing their order, before returning to their conversation. Social actors can and often do have multiple loci of attention, with each action still located somewhere on the attention/awareness continuum.

Using this model, attention/awareness is determined analytically based on both the complexity of a higher-level action, and the intensity of the actions that constitute it. This compound notion is called modal density, and the higher the modal density of an action, the higher its level of attention/awareness. In our example of ordering coffee, complexity refers to the multiplicity of lower-level actions directed towards conversation, and towards ordering coffee. The two people we observe lining up for coffee orientate their postures towards each other: they look at each other, and speak to each other. These actions demonstrate their attention towards conversation. At the same time, they are in a line moving towards the coffee counter. Their participation in this line, through posture, layout and an occasional gaze shift demonstrates their attention towards lining up for coffee. The complexity of each higher-level action (conversation, ordering coffee) can be used to begin the analysis of attention/awareness for each social actor.

In addition to the complexity of each higher-level action, we must also analyse the intensity with which lower-level actions are produced. Highly intense actions demonstrate a high level of attention/awareness. The intensity of actions refers to the strength with which an action is produced. This notion of strength is relative to the interactional environment. In a recent paper, Fogarty-Bourget, Pirini, and Artemeva (Forthcoming) define modal intensity as the:

phenomenon of heightened pragmatic salience occurring as a result of an action being made prominent relative to surrounding or co-occurring actions that have become recognized as regular, expected, or standard due to frequency of occurrence.

(Fogarty-Bourget et al., Forthcoming, p. 4)

Actions can take on high modal intensity through many different qualitative features. Fogarty-Bourget et al. (forthcoming) identify features such as distribution, size, pitch and visual/textual devices. In the case of ordering coffee, as someone pauses their conversation to stare at the order board, they produce gaze and posture with high intensity, indicating their focus on ordering coffee.

Importantly, we address interactional attention/awareness. ‘Interactional’ refers to the attention/awareness that we can determine from the actions people produce. Of course, there is a link between what people appear to be attending to, and how they experience their attention/awareness. As social actors we make sense of the social world by judging what people appear to be attending to, and what they appear to be aware of. We also demonstrate our own attention/awareness to others. These are core processes for interacting with others, and we utilise these as analysts when we conduct a Multimodal (Inter)action Analysis of attention/awareness.
However, as analysts it is important to distinguish between what people demonstrate interac-
tional attention/awareness of, and what they are experiencing. People are clearly capable of
misdirecting others on purpose, or inadvertently, when it comes to their attention.

**The foreground/background continuum of attention/awareness**

The foreground/background continuum of attention/awareness positions higher-level actions
along a continuum of attention/awareness from the focus, or foreground, through the midground
to the background. Higher-level actions are positioned based on modal density. The higher the
modal density of a higher-level action, the further to the left it sits on the continuum.

Returning to ordering coffee, the conversation with the friend sits at the foreground of
attention indicated by modal density (see Figure 40.1). Of course, neither friend forgets that
they are ordering coffee, but they are able to attend to their coffee order with a low level of
attention/awareness. This lower level of attention/awareness is demonstrated through a lower
density of actions directed towards ordering coffee.

Analytically determining the attention/awareness of social actors is central to the analysis
of multimodal interaction. Norris (2019) points out that many traditions including conversa-
tion analysis and discourse analysis make assumptions about a link between focused attention
and talk. However, while people often focus on their talk, this is not always the case. Under-
standing social interaction requires an analytical approach to interaction which is capable of
assessing the way that social actors shift their focused attention, and remain aware of actions
in the mid- and background.

**Site of engagement**

The site of engagement is an important conceptual tool allowing analysts to move between micro,
intermediate and macro analyses (Norris, 2011; Norris & Pirini, 2017). The notion was first intro-
duced by Scollon (1998) as the “window opened up through the intersection of social practices”

![Graph of attention/awareness](adapted from Norris, 2004)
(p. 11) and later refined by Norris and Jones (2005) as “the ‘real time window’ opened through the intersection of social practices and mediational means that enables a mediated action to occur” (p. 139). As a real-time moment through which mediated action occurs, the site of engagement links social practices, discourses and mediational means. The concept itself acknowledges the one-time, unique and unfinalizable nature of mediated actions occurring in the world while acknowledging how these one-time unique actions link to practices. Analysts are thus able to move between different levels of understanding, from the micro-level actions produced in concrete moments, to larger-scale practices and discourses intersecting within a site of engagement.

**Scales of action**

Multiple higher-level actions can be produced at one time, but they can also be produced on different timescales. Scales of action as a methodological tool facilitates the analysis of how chronologically unfolding lower-level actions can actually produce different higher-level actions on various timescales.

Norris (2017) details the analysis of scales of action through a data corpus involving interactions while driving. Despite the sequential unfolding of utterances during the drive, the actions build various higher-level actions on different timescales. The conversation which unfolds during the drive may turn to a debate about political leanings pertaining to current electoral candidates, and this debate could have began hours prior to the drive. In this case, the actions unfolding during the drive do not simply sequentially construct a conversation during the drive, but consequentially build and are influenced by a larger-scale action initiated well before the drive and persisting long after its completion.

The drive itself and the ongoing conversation are also embedded in the action of shopping for art supplies. Multiple other actions unfold as well, like opening and closing the door or reading of a letter. A sequential approach to these actions would see them as integral in the unfolding of the conversation but perhaps miss their relationship to larger-scale actions being performed on larger timescales. The reading of the letter, while unfolding sequentially, is not integral to or related to the action of shopping for art supplies but rather to the larger-scale action of collecting child support. This notion of scales of action facilitates the analysis of the complex interrelationships of actions which unfold in real time and their relationship to larger scales of action. This helps the analyst tease apart the complexities of interweaving higher-level actions which are continuously produced across varying timescales.

**Insights generated from employing Multimodal (Inter)action analysis**

To date, innovations around Multimodal (Inter)action Analysis have primarily been directed towards improving or enabling analysis of multimodal action. More recent work uses the framework to open up new directions for applied linguistics research. In this section we highlight two recent areas of development enabled through the theoretical and methodological aspects of the framework previously highlighted. The first relates to intersubjectivity, and the second relates to the development of communicative competence.

**Intersubjectivity in knowledge communication**

Intersubjectivity is typically treated as shared understanding, arrived at through talk. Indeed, work in applied linguistics that takes a sociocultural approach focuses on language as the
“most pervasive and powerful cultural artifact that humans possess to mediate their connection to the world” (Lantolf & Thorne, 2006, p. 201). However, as we pointed out previously, prioritising language prior to actual analysis is problematic.

A multimodal approach to intersubjectivity reveals different tiers of material intersubjectivity (Pirini, 2016). Actions like producing posture and using layout are materially persistent, since people typically hold a posture for a while or use the layout of a space to sit in a chair or lean against a bar. These actions produce a materially persistent substrate, which is required for (inter)actions involving actions that exhibit a more fleeting materiality, such as gesture, gaze and spoken language. Without a somewhat stable material alignment, people cannot hear what other people say, or see gestures, head movements and so forth. Pirini (2016) identifies three tiers of material intersubjectivity. The first, most durable, tier typically establishes stable proxemic relationships through layout and use of furniture. The second tier is made up of actions that persist, but are adjusted during interaction like postural shifts. These durable tiers make up an interactive substrate upon which more fleeting actions are produced. The tier which exhibits the most fleeting materiality involves spoken language, gesture and gaze.

Including all modes and their materiality allows analysts to explore how features of the materially persistent environment impact upon shared understanding. One example of this approach comes from an ongoing study into video conferencing. The camera is ever-present in these interactions, and participants have learned to orientate to the camera to display their attention and participation, including visually introducing new topics by introducing objects (Geenen, 2017). The camera becomes a central material feature in the higher-level actions that social actors co-produce. Another recent paper (Pirini & Geenen, 2018) shows how participants prioritise shared gaze orientation prior to establishing verbal alignment.

Intersubjectivity from this perspective is taken as a materially constituted (somewhat) shared experience of the world. At a more abstract level there is overlap between practices and discourses reverberating through the site of engagement, which builds on the materially constituted overlap. The precise make up of this materially constituted shared experience has effects on patterns of interaction. It is the interplay between the constituents of shared (multimodal) experience and how social actors produce meaning that most likely offer rich grounds for further scholarship in areas traditionally explored in applied linguistics, such as language acquisition.

**Developing communicative competence**

Multimodal (Inter)action Analysis has proven particularly useful for understanding the emergence of pragmatic aptitudes and how they are acquired through real-time social interaction. Non-verbal modes of communication have always figured prominently in developmental pragmatics, however, empirical concern has traditionally centred on the pre-verbal period of communicative maturation. Recently, efforts have been made to unravel the extent to which non-verbal modes may figure in pragmatic development alongside language (Clark, 2014; Geenen, 2018, forthcoming; Norris, 2019).

The inclusion of all communicative modes in research has revealed that very young children (2 years) are sensitive to subtle alterations in participation frameworks (Geenen, 2018). They appear adept at tracking conversational topio and re-introducing said topics to newcomers who may have ‘missed something’. The actions themselves, however, are more nuanced and do not always occur through the mode of spoken language as children draw upon a range of non-verbal modes like gesture and object handling to introduce topics (Geenen, 2017, 2018). Only a methodology which prioritises the *pragmatics of interaction* (Clark, 2014) can help accurately deduce when and to what extent communicative aptitudes emerge. Traditional
linguistic methods which uniformly focus on language production run the risk of conflating language use and communicative aptitudes or even implicitly suggesting socio-cognitive comorbidity for language specific delays in development.

Using Multimodal (Inter)action Analysis to inductively tease apart the complexity of peer-child multi-party interactions has revealed that more mature conversationalists frequently defer interactional agency to younger children. Adult peers use tangible objects in the material surround as a means to engage children. Importantly, peers use the physical objects in the here-and-now as a bridging mechanism to experiences then-and-there (Geenen, forthcoming). Encouragements to undertake non-verbal communicative actions with tangible objects link the actions and objects to experiences in the past. This motivates developing interlocutors to consider whether personal experiences which occurred in the past might be relevant conversational topoi, therein motivating explicit judgment about conversational relevance. These strategies support episodic retrospection and make tangible the link between the current interaction and past experiences.

**Conclusion**

In this chapter, we have introduced the central theoretical tenants which inform Multimodal (Inter)action Analysis as a methodological framework for the qualitative analysis of real-time social interaction. The focus on mediated action as a starting point keeps alive the complex individual, sociocultural, institutional and historical components of human action. This recognition of the social and historical in discrete forms of human action ensures that analysis implicitly recognizes the irreducible tension between the individual and society. Action does not occur without mediation; thus, contestation regarding theoretical or analytical primacy is overcome by maintaining the mutual influence of both individual and society in a single analytical unit.

Additionally, Multimodal (Inter)action Analysis explicitly accommodates for the varying materiality and structural organisation of communicative modes while maintaining a single unit of analysis. Action as an entry point to analysis also facilitates an approach to the physical environment and objects within it as themselves constituting frozen actions – as the material remnants of humans interacting with and through their environment.

As multiple complex chains of lower-level actions undertaken through various communicative modes are integral in the building of higher-level actions, charting modal configurations enables the analyst to discern the hierarchical salience of any communicative mode in the construction of that higher-level action. Through teasing apart the relationship of individual modes and their pragmatic function to the higher-level action, analysis can capture how and when certain modes take on communicative saliency within the interaction. Attendance to individual modes and their configurations can also help analysts discern which higher-level action is in the foreground of a social actor’s attention/awareness continuum. As people can and often do perform multiple higher-level actions simultaneously, attending to modal complexity and intensity can assist in determining how and when particular higher-level actions move in and out of a social actor’s focused attention during any period of social interaction.

We have attempted to provide a functional overview of the methodological framework, though admittedly certain concepts and tools have been omitted in an attempt to more thoroughly present and discuss others in the framework. One such omission has been the analytical protocol and development of multimodal transcripts which go hand-in-hand in the analytical procedure. For a complete and comprehensive guide to analytical protocol and transcript generation, see Norris (2018). We could not deal comprehensively with topics like the mediational
interrelationship (Geenen, 2013b, 2014), the notion of practice (Norris, 2004, 2011; Scollon, 2001) or other developments, such as what constitutes a communicative mode and how they develop and are during ontogenesis (see Norris, 2013).

An analytical focus on action has numerous theoretical and methodological advantages when analysing real-time social interaction. As both a methodological starting point and a theoretically grounded concept, the mediated action ensures that analysis itself maintains the anti-reductionist ethos championed in the socio-historical psychological approach to human psychological functioning popularised by Vygotsky. The methodological framework can be utilised across a diversity of thematic domains and is perfectly suited to research questions and hypotheses which have an empirical focus on the complexities of social interactions, which are always and undeniably multimodal in nature.

References
Fogarty-Bourget, C., Pirini, J., & Artemeva, N. (Forthcoming). The role of modal intensity in promoting students’ involvement in learning.
Geenen, J. (Forthcoming). Objects and materiality in pragmatic development: Here-and-now to then-and-there.