Interactive repair and the foundations of language

Mark Dingemanse and N.J. Enfield

The robustness and flexibility of human language is underpinned by a machinery of interactive repair. Repair is deeply intertwined with two core properties of human language: reflexivity (it can communicate about itself) and accountability (it is used to publicly enforce social norms). We review empirical and theoretical advances from across the cognitive sciences that mark interactive repair as a domain of pragmatic universals, a key place to study metacognition in interaction, and a system that enables collective computation. This provides novel insights into the role of repair in comparative cognition, language development, and human–computer interaction. As an always-available fallback option and an infrastructure for negotiating social commitments, interactive repair is foundational to the resilience, complexity, and flexibility of human language.

An inversion of perspective

Culture, communication, and cognition are enabled by the extraordinarily complex and generative systems of language [1–3]. Yet, at first sight, language appears far from perfect. In everyday language use, errors (see Glossary), misunderstandings, and repairs are frequent, occurring every few turns [4,5]. Linguists and cognitive scientists have long treated these features of social interaction as a distraction: the rubble of performance from under which we must excavate pristine competence. But recent work from across the cognitive sciences suggests an inversion of perspective. This work finds interactive repair to be an integral part of the language system and its social-interactional infrastructure. Interactive repair is therefore foundational to the flexibility and resilience of language.

The findings we review here are emerging from diverse work across the cognitive and behavioural sciences, including interactional linguistics, psycholinguistics, the sociology of action, experimental semiotics, language development, and comparative ethology. Interactive repair provides a common focus for disparate strands of work that together shed light on the interactional foundations of human social cognition and communication [6,7].

But interactive repair is more than a reactive quality-control system. It offers people opportunities at every turn to review, revise, and recalibrate around meaning in real time, editing not just the flow of information but the interpersonal commitments that the exchange of information creates. Acknowledging these two core functions of interactive repair – as a mechanism ensuring informational robustness and as an organisation for social accountability – will improve our understanding of language and mind.

Interactive repair: what it is and where to find it

Repair is relevant to any field touching on human interaction, and, accordingly, it has a rich history with considerable terminological diversity [6]. Interactive repair, the term we use here, highlights the role of both parties in reaching and calibrating mutual understanding and is relatively agnostic.
about responsibility for trouble. It is more specific than a broader notion like feedback [8,9], more neutral than terms that imply faults (e.g., disfluency [10,11] or error [12]), and more encompassing than terms that foreground only one element of the sequence (e.g., breakdown [13,14] or clarification request [15,16]).

A commonly made distinction is between self-repair and interactive repair. Self-repair has been widely studied in psycholinguistics, mainly on account of its high frequency in spontaneous monologic speech [17,18]. In self-repair, the producer of an utterance makes an error and immediately corrects it themselves. This may include errors of articulation, grammatical errors, lexical errors of word selection, and pragmatic errors of appropriateness. Self-repair may sometimes be interactionally contingent, such as when someone producing a turn sees that a recipient is puzzled and revises their utterance midstream [19,20]. Interactive repair – also known as ‘other-initiated repair’ and most specifically as ‘other-initiated self-repair’ [21,22] – involves an explicit exchange between two participants to identify and correct a problem, calibrating mutual understanding across turns. This is the form of repair in focus here.

Turn-taking and repair: interdependent systems

To study interactive repair, one is necessarily studying interactional sequences, often referred to using the term ‘turn-taking’. Turn-taking and repair have been co-implicated from the earliest account of the generative rules of spontaneous conversation: ‘the turn-taking system lends itself to, and incorporates devices for, repair of its troubles; and the turn-taking system is a basic organizational device for the repair of any other troubles in conversation’ [23]. This explicit recognition of the possibility and need for repair in human turn-taking was an important advance on earlier models that did not discuss trouble and its resolution in the flow of conversation [24].

A sequence of interactive repair has three parts (Figure 1A, Key figure): the pivot is a repair initiation, pointing back to a trouble source and forward to a repair solution [21]. Repair initiations can come alone or in bursts if turbulence requires multiple rounds of initiation and attempted clarification [25]. Normally, a move in conversation should ‘fit’ to the prior move, building a sequence of action: a question invites an answer, a greeting a countergreeting, and storytelling invites continuers like mm-hm” [26]. Although a fitted response may be interactionally preferred, you can only produce one if you have grasped the prior move. Repair provides a fallback option that is always available in case the prior move was not heard or understood (Figure 1B).

Repair initiations come in three types based on how they target the prior turn and the kind of response they invite in the next [27]. In pointing back to a prior turn, a repair initiation may leave open what or where the problem is, or it may restrict the search space in some way; in terms of inviting a response, it may request repetition, clarification, or elaboration, or it may offer a possible solution to be confirmed or disconfirmed. These operations define the three basic repair initiation types of open request, restricted request, and restricted offer (Figure 1C). Again, these options embody an asymmetrical system of response types: producing a restricted format such as ‘who?’ requires some grasp of the prior move, whereas an open format like ‘Huh?’ is a fallback option. Across all known human languages (signed and spoken), this same system of interactive repair offers the means to manipulate utterances and make visible their internal structure [27–29].

The nested asymmetries of the repair machinery lend it a certain interactional logic. Thus, to produce a continuers or a fitted response is to forgo the opportunity of initiating repair on the prior turn (and so to imply that it was understood) – often the preferred route [22]. Conversely, within repair,
to choose an open repair format is to display no grasp whatsoever of the prior turn, which may be interpreted as interactionally uncooperative or disruptive.

Repair carries a significant functional load in human interaction. Even when counting only independent repair sequences, repair is frequent: Rarely do 4 min go by in conversation without an interactive repair sequence occurring (Figure 1D). This is not surprising, given the infinite flexibility – and thus low predictability – of human language [2,3] combined with the intense time pressures on production and comprehension in the context of turn-taking [30,31]. Moves and countermoves can be entirely novel and must be processed on the spot. This processing must happen very fast: when it comes to the timing of turn delivery, language users are sensitive to the pragmatic import of differences on a subsecond timescale [32]. All this means that our use of language is highly susceptible to errors and noise, with many failure modes potentially resulting from this.

The sheer frequency of interactive repair demonstrates its constant relevance to language in interaction. Its collaborative nature aligns with the nature of conversation as joint action. Its metalinguistic operations on the linguistic system embody the unique reflexivity of human language.

Having repair is vital, but using it is not without cost [33,34], and so, although it is frequently used, there are also ways of avoiding it, for instance by simply not initiating repair (even when needed) or by providing noncommittal responses [35–37]. In this way, interactive repair provides the language system with a form of organisational slack [38]: a cushion of resources available for

---

**Key figure**

Anatomy and frequency of interactive repair

A Interactive repair (other-initiated)

<table>
<thead>
<tr>
<th>Trouble source</th>
<th>Repair solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self</td>
<td>Oh Sibbie’s sistuh hada baby go:way</td>
</tr>
<tr>
<td>Other</td>
<td>Sibbie’s sister.</td>
</tr>
</tbody>
</table>

B Response options after someone’s turn

- Fitted response (preferred)
- Interactive repair (always an option)
- Restricted (preferred)
- Open (always an option)

C Repair formats as a window onto turn structure and syntax

- Restricted Offer
- Restricted Request
- Open Request

D Frequency of interactive repair

- Mode: 17 s
- Average: 84 s
- 95% of cases: < 248 s

---

Figure 1. (A) With interactive repair, another participant initiates repair, inviting a repair solution by the first; the repair initiation is a pivot, pointing both back and forward. (B) While a fitted response is preferred, initiating repair is always a possible next move; likewise, within repair, while a restricted format is preferred, an open format is always an option. (C) Across diverse languages, formats for interactive repair range fall into three types, depending on how they target the trouble in prior turn and the kind of response they typically invite; these can be ranked from less to more specific in terms of the grasp of the trouble source they display. (D) Empirical cumulative distribution of independent repair sequences (black curve) as they occur over time in informal conversation in a global sample of 12 languages (grey curves). Across languages, the steepest part of the slope is around 17 s, the average 84 s, and nearly all sequences occur within a 4-min window from the last (data from [4]).
discretionary use that lends resilience to the larger system. Exactly where the balance lies between initiating repair too much and not enough is an empirical question. Answering this question is challenging because it would require knowing when someone might have repaired but did not, though it may be possible to derive information-theoretic predictions from models of repair. There is a key role here for empirical work on repair in special settings, for example with people who are hard of hearing or have other communication difficulties [39–41] (Box 1).

Core properties of language: reflexivity and joint accountability
So far, we have reviewed structural features of interactive repair along with its functions as an informational safeguard in the context of language’s high complexity and low predictability. We now consider two further features of interactive repair, which are also core properties of language uniquely elaborated in human interaction. These features are reflexivity of signalling and normative accountability of joint commitment. Language would be impossible without these, and interactive repair is a key mechanism through which they are realised.

Reflexivity
A defining feature of interactive repair in human language is its reflexive character. Utterances like ‘What did you say?’ presuppose the capacity to refer to what was said. Language appears to be the only communication system that can communicate about itself in this way. As the linguist Charles Hockett remarked in listing the design features of language: ‘Bees dance about sites, but they cannot dance about dancing’ [42]. The consequences of this are underappreciated [43]. Without linguistic reflexivity, we could not define words or ask what a word means, written language would not exist, there would be no way to report others’ speech or pass on messages, and there would be no interactive repair [44–46]. Species unable to signal about their own signalling systems have no means to secure joint attention on their interactions as they occur and thus have no means to convey messages like ‘Huh?’ or ‘Did you mean X?’.

Box 1. Advances in conversation analytic work on repair
The interactive repair system was first described by conversation analysts in the 1970s [21,122] and has recently seen a flowering of empirical work across a wide range of languages and interactive settings. Here, we highlight three strands of work that bring out the double-edged nature of interactive repair as a system for the public negotiation of understanding and a site where misunderstandings are brought to the surface.

1. First, cross-linguistic comparative work proposing universal principles of interactive repair [4] has been replicated and refined in a growing number of spoken and signed languages around the world [28,123–125]. Related lines of research trace the repair system in specialised settings beyond informal conversation. For instance, interactive repair is increasingly recognised as a key aspect of interactional competence in (second) language learning and in classroom settings [126–128].

2. Work on repair in safety-critical settings has brought to light the limitations of stipulated communication protocols, and the ways in which people use interactive repair to creatively work around them, for instance in ship-to-ship communication during icebreaker operations [129] or in radio telephony during military patrolling exercises [130].

3. Second, there is increasing attention to multimodal aspects of repair, from gaze patterns and facial signals [131,132] to upper body posture and manual gestures [133,134], which may accompany or even replace more conventionalised lexical resources. This work shows that people recruit all semiotic resources at hand to signal degrees of understanding and finds that principles of minimising collaborative effort extend to how people combine gestural and lexical elements [95].

4. Third, work on repair with clinical populations has shown how the repair system is mobilised in interaction that involves people with disabilities [41]. This work has found that interactive repair can facilitate the rapid calibration of mutual understanding [135,136] but also that people can treat it as a possible source of turbulence to avoid [37,40]; a balancing act that is perhaps most sharply felt in highly asymmetrical settings and that is in need of more empirical work.

A fundamental challenge in cognitive science is that we lack direct access to cognitive processes. As is evident in conversation analytic work on repair, this lack of access is as much a problem for participants as for analysts, and the public displays of understanding we see in interaction provide a key part of the solution. Interactive repair sequences provide one of the richest sites to study processes of mutual understanding. Like an exploded view diagram, they make visible the separate elements of successful communication and how they are distributed over minds and across turns in interaction.
There is a chicken-and-egg quality to the observation that interactive repair both exploits and embodies the reflexivity of language. This calls for a dialectical approach [47,48]: Repair and reflexivity are in a mutually reinforcing feedback loop, with repair relying on reflexivity and at the same time being the engine by which reflexivity gains much of its adaptive value. The reflexive function of language, as embodied and exploited in interactive repair, also enables a second core property of the uniquely human communication system: namely, mutual and public accountability within the context of joint action and shared intentionality.

**Joint accountability**

Using language requires joint commitment in the strong sense, which means not only that participants in interaction are working together but also that their commitment to doing so is subject to public attention and scrutiny. In human interaction, each move is potentially subject to praise, blame, rebuke, negotiation, checking, double-checking, and more [49,50]. These operations are typically carried out using language and, in human interaction, are often directed towards language itself [51–53]. A joint commitment not only ‘obligates the parties one to the other to act in accordance with the commitment’ but also bestows parties with the authority ‘to demand conforming action and rebuke for non-conformity’ [50]. This authority or right to enforce norms and to sanction departures is accompanied by a participant’s duty to produce appropriate actions and to expect rebuke for departures. A corollary of this duty is that participants should preemptively account for failures and departures when they occur (hence the term ‘accountability’). People will announce or explain if they need to disengage from a joint activity [54].

Linguistic interaction is subject to this regime of joint accountability [7]. This applies to both roles that participants play: producer and recipient (we use these terms as modality-agnostic variants of speaker and listener). Just as a producer is accountable for doing their part – for communicating in ways that are relevant, understandable, and perceivable – a recipient is accountable for ensuring that they have perceived and understood what was said to them. This is not (only) a matter of ‘politeness’ but a condition for interaction being possible at all. Interactive repair comes to the fore when there is a failure on either side: whether my saying ‘Huh?’ points to a problem of your speech or of my hearing, it addresses both parties’ joint commitment to keep the interaction on track.

Thus, through the use of interactive repair, dialogue is regulated by social norms, which provide ‘stable expectations for navigating our social world’ [55]. Such expectations, often unspoken, can be defined by two qualities of normative behaviour: (i) the behaviour is not noticed or commented on when the norm is followed (subliminal), and (ii) it is noticed and commented on when the norm is transgressed (abliminal) [7]. Repair plays a key role here in that it surfaces precisely when there are errors, disruptions, or transgressions. Although the idea of norms for successful social coordination often invokes simple arrangements such as which side of the road to drive on [55], language is a special case because the failure modes are of a different nature. Driving on the wrong side of the road risks potentially deadly consequences, but with language, failure might merely mean interactional turbulence or, more rarely, breakdown of a conversation [21]. While norms can achieve convergence through coordination, they cannot succeed without an always-available system for holding participants accountable for upholding their part of the bargain [56–58].

Further emphasising the relation between interactive repair and norms, while repair is often seen as a mechanism for dealing with informational uncertainties [36,59], the failures being repaired (or norms being enforced) may also be matters of appropriateness [60,61]. In one example, when a child says, ‘Put on the light’, the mother responds with the repair-initiator ‘Pardon?’ This is repaired
as ‘Put on the light please’, and the mother says ‘Better’ before putting on the light [62]. Here, the repair targets a norm of politeness. In another kind of example, when there is unclarity about the reason for a practical request (someone asks for a spoon when they already have one), rather than directly questioning why the request is made, one can instead issue a repair initiation, allowing the other to rephrase the request and add a reason (e.g., the spoon they have is dirty) [63]. These observations show that repair is not merely a way of dealing with informational failure but is just as much a system for administering social accountability.

**Biological precursors for interactional repair in humans**

Despite the centrality of repair in the received model of turn-taking in humans, literature on turn-taking in non-human animals seldom mentions repair at all. This makes sense when turn-taking refers simply to alternating signals that appear to avoid overlap (i.e., turn-taking in a broad sense, or TTB). By contrast, turn-taking in a narrow sense (TTN) is built out of language and could not function without a reflexive repair system (Box 2).

The difference between TTB and TTN is anything but trivial. The average linguistic system provides thousands of words, idioms, and morphosyntactic rules that give language its uniquely generative character [2,3,64]. No other animal communication system is known to build moves using symbols plus generative morphosyntax (i.e., words and rules for combining them). Instead, many such systems draw on relatively limited inventories of signals in exchanges that lack most or all of the flexibility, accountability, and reflexivity of human interaction. Non-human interaction does not appear to show the kind of open-ended interactive repair

---

**Box 2. Turn-taking: broad and narrow senses**

Given the growing interest in turn-taking and its cognitive underpinnings in our species, it is remarkable that comparative work on turn-taking so rarely mentions or accounts for interactive repair (but see [65]). We explain this by distinguishing between two notions of turn-taking (Table I). Recent work on animal interaction uses the term ‘turn-taking’ to refer to alternation of signalling among conspecifics where interactants avoid overlap. This can be termed ‘turn-taking in the broad sense’ (TTB). TTB is observed in a wide variety of species [137] and even in interspecies interactions [138]. On this broad view of turn-taking, the phenomenon is evidenced across all the major branches of the primate order [31], suggesting homologs among human and non-human systems of interaction. Some discussions of interactional sequencing among animals loosely mention turn-taking ‘rules’ [139,140], but this relates mostly to the tendency to avoid overlap and seldom considers the explicit defining rules of human turn-taking discovered and described in the field of conversation analysis [23].

By contrast, turn-taking in the narrow sense (TTN) has more specific properties, the most important of which relate to flexibility in recipient selection, unfixed and recursively complex internal structure of turns, and the possibility of carrying out interactive repair across turns. (We note that there is a distinction between turn-taking and sequence organisation [141], but we set it aside here.) Without interactive repair, there would be little adaptive value in the other features. After all, the more complex a communication system becomes, the more sensitive it is to perturbations. The more ways there are to say something, the more ways there are to be misunderstood. The distinction between narrow and broad senses of turn-taking helps to explain why repair has flown under the radar in most studies of animal communication outside humans. Only turn-taking in the narrow sense offers the requisite complexity and the normative accountability that renders interactive repair both possible and necessary.

**Table I. Turn-taking in the broad sense and in the narrow sense**

<table>
<thead>
<tr>
<th>Property</th>
<th>TTB</th>
<th>TTN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternating signals</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Avoidance of overlap</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Flexible recipient selection</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>Recursively complex turn structure</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>Possibility of carrying out interactive repair</td>
<td>x</td>
<td>✓</td>
</tr>
</tbody>
</table>

aAbbreviations: TTB, turn-taking in the broad sense; TTN, turn-taking in the narrow sense.
seen in human interaction [65]. Some linguists have singled out the morphosyntactic
generativity of human language as the magic bullet that answers the question ‘why only us?’
[66]. Here, by contrast, we draw attention to the evolved interactional infrastructure necessary
to maintain a communication system that is so radically compositional and creative. There
would be little adaptive value in an infinitely generative system without interactive repair to de-
fend its flexibility and bring needed robustness to the system.

New cross-species work on comparative cognition [67–69] suggests an evolutionary continuum
that could underpin the human interactive repair system, building on elements that are present in
species such as other great apes [65,70]. There is evidence that animals engage in self- and
other-monitoring and that they persist with, or elaborate, communicative actions that have failed
to elicit a relevant or desired response. Non-human animals may show confusion at unexpected
actions or persistence in the face of failure, but they do not hold each other normatively to account
for errors of production or comprehension of actions: that is, they may issue a sanction in some
physical form, but without language they seem to be unable to isolate or describe the transgres-
sion. Joint commitment in the strongest sense requires public social signalling about the joint
commitment. This is indeed a prerequisite for interactive repair, where the joint commitment
at hand is the cooperative use of language. Only in human interaction, it seems, do we see explicit
metacommunicative operations that exploit the reflexivity of language and reveal joint commit-
ment in the strongest sense [71,72].

In sum, language-mediated human interaction shows three key properties. These are (i) lexicon
and generative morphosyntax, (ii) reflexivity of signalling, and (iii) joint accountability of interactional
agents. Taken together, these properties are the ingredients for a system of interactive repair
such as we observe wherever language is used.

Consequences of interactive repair
Interactive repair not only is an important domain of study in its own right but also is key to numerous
areas of research in the language sciences. Here we focus on three areas: language development,
language processing, and conversational interfaces.

Language development: repair as a metalinguistic scaffold
According to one estimate, as many as 50% of infant-initiated interactions involve some form of
interactive repair [73], making repair a key feature of interactional infrastructure in early language
development [74–76]. The child encounters repair alongside its earliest communicative moves,
which are repeated, recast, and recontextualised by caregivers to weave webs of social partic-
ipation [77].

Consider two observed examples (Figure 2). By asking ‘What did you say?’, a caregiver does not
merely get the child to repeat their turn; the caregiver also conveys to the child that the child is
accountable for speaking clearly and understandably (Figure 2A). By reformulating a prior turn

A  Margaret brings cup to Father in kitchen

M  ((indecipherable))
F  What did you say?
M  Want more milk!
F  Mamma said that was enough.
M  No! More!

B  Margaret stands in front of DVD player

F  You mean you want to see the kitty-cat movie?
M  Yeah, yeah.
F  Okay, but you just watched it this morning.

Figure 2. Interactive repair sequences in language development. Adapted, with permission, from [77]. Abbreviations:
F, father; M, Margaret.
prefaced by ‘You mean’, a caregiver demonstrates how people can ascribe detailed intent to simple phrases and how such simple phrases can be expanded to more grammatically complex formulations (Figure 2B). Closing the interactive repair sequence can be done either by directly returning to the base sequence (as in Figure 2A) or with a sequence-closing ‘Okay’ (as in Figure 2B); both methods model for the learner how interactive sequences can be nested and how, after completing the repair side sequence, one returns to the business of the base sequence.

Repair sequences often segment utterances at constituent boundaries and streamline the querying of some elements (e.g., standalone words) more than others (e.g., bound morphemes), reflecting grammatical asymmetries. By making morphosyntax visible (Figure 1D) and showing how simple phrases can be paraphrased or expanded (Figure 3B), interactive repair displays the compositional nature of linguistic systems. Thus, far from being a mere remedial procedure, interactive repair helps a child build lexical knowledge (by bringing about the repetition of words and phrases), pragmatic skills (by showing how meaning is negotiated in context), and grammatical insight (by profiling morphosyntactic structure) [78,79].

More speculatively, the tremendous frequency and the overt metalinguistic nature of repair may make it a key locus for early language learners to understand the mechanisms of reflexivity and social accountability. Tracing the development of interactive repair and its consequences in language development and socialisation is a key area for future research [15,76] (see Outstanding questions).

Planning and prediction: repair as collective computation

Human interaction is never a one-shot game [80], because interactive repair is available when needed. This has important consequences for the dynamics of the language system, which come into view when we shift from a microgenetic frame (focused on individual language processing) to an enchronic one (focused on turn-by-turn dialogue). In the enchronic frame, the
availability of interactive repair considerably relaxes the ‘now-or-never bottleneck’ in language processing [81] by enabling constant opportunities for revision. As an always-available safety net that makes ambiguity manageable and good-enough processing viable [82,83], interactive repair is a key way in which people share computation.

Any linguistic interaction is a coordination problem [84,85]: We cannot read each other’s minds; yet, we have to coordinate our actions and calibrate our understandings [86,87]. The question of how people come to the needed kind of distributed cognition can be tested using computational modelling, as suggested by a flourishing literature on language games and pragmatic reasoning in experimental settings [36,88,89]. Pragmatic reasoning can quickly become computationally complex because it requires weighing myriad hypotheses about what the other might think that you might think and so forth, recursively [90]. In state-of-the-art models of pragmatic reasoning, larger lexicon sizes create a combinatorial explosion of possibilities and attendant computational costs [91].

Interactive repair offers a partial solution. When people engage in interactive repair, they distribute computational and representational work over multiple minds and turns: a form of distributed cognition [92] that we can also think of as collaborative computation [93,94]. Agent-based modelling shows that a simple form of interactive repair can dramatically alleviate the computational complexity associated with larger lexicon sizes (Figure 3, based on [91]). Two kinds of agents are compared: (i) reasoners (pragmatic reasoning-capable agents without repair) and (ii) repairers (repair-capable agents without pragmatic reasoning). As lexicons grow, reasoners suffer decreasing communicative success, whereas repairers’ rate of success is hardly affected (Figure 3A); in addition, computational complexity increases quadratically for reasoners but only linearly for repairers (Figure 3B). Thus, interactive agents equipped with a simple form of repair and the possibility of taking multiple turns show superior communicative success with lower computational cost at larger lexicon sizes.

Of course, there is a cost to repair. Observational studies show that people work to divide this cost in a way that is most effective for the dyad [4,95], and they use their successive turns to rapidly narrow down problems and resume the interaction [4,25]. The modelling work reviewed here [91] reveals a similar pattern (Figure 3C): When needed, repair-capable agents take additional turns to reduce entropy (uncertainty about meaning) in cost-effective ways. In all, this modelling work suggests that a simple form of interactive repair can be a significantly less computationally costly strategy than pragmatic reasoning for securing communicative success.

Clearly, people can use both pragmatic reasoning and interactive repair, and a key question for further research is how they choose between the two or when and how they combine them. Modelling helps us better understand the relationship between cognitive processes and interactional resources and thereby can guide further observational work, inform experimental design, and support theory building [96,97].

Conversational interfaces: repair as a new frontier for mutual understanding and accountability
Artificial interactive interfaces and conversational agents are taking on ever-larger roles in everyday life, but they tend to support only limited dialog flows, cannot gracefully repair breakdowns, and do not partake in the social commitments and accountability that animate human interaction [13,98]. These constraints are, of course, closely related. Present-day conversational interfaces excel at prediction and emulation but lack the freedom and flexibility supported by a sophisticated repair system [99,100], adversely impacting user experience [13,101,102].
Repair strategies beyond the simple open request (‘Huh? Could you repeat that?’) are notoriously hard to implement in dialog flows due to the combinatorial explosion of possibilities in free-flowing interaction: any element of a prior turn may or may not be a cause of trouble [99], which makes formulating transparent repair initiations and solutions a formidable problem [103,104]. For people, knowing what you do not know and coping with ambiguity are key metacognitive feats scaffolded by lifelong experience [105]. In contrast, today’s conversational agents are fine-tuned to come across as confident despite lack of understanding [106,107], and only when faced with dead ends in dialog flows will they attempt simple forms of repair [13].

There are two fundamental challenges for conversational interfaces today, and repair is tightly interlinked with both. The first is the nature of understanding. Text generators based on next token prediction completely lack understanding [108]. Even as people may be impressed by the fluidity and apparent interactivity of such interfaces, the division of labour remains highly skewed, with the bulk of the interpretive and interactive work falling to the human [109,110]. This is best seen in the applied field of prompt engineering, the iterative refinement of textual prompts to coax a text generator towards desired responses [111]. Empirical work is needed here to test the lengths people are willing to go to in order to achieve a semblance of understanding. For instance, whereas people rarely escalate repair sequences beyond three or four rounds, it is an open question how this affects people’s sense of joint commitment in interactions with machines.

The second challenge relates to social accountability. Whether following predesigned dialog flows or generated by next token prediction, the output of conversational interfaces is fundamentally devoid of accountability: it is text produced without a commitment to truth or joint action [72]. By tailoring textual output to be maximally in line with what human raters prefer, techniques like reinforcement learning with human feedback obscure this lack of accountability but do not fundamentally change it: a text generator may ‘apologise’ when prompted for clarification or correction but will indifferently regurgitate the same misunderstanding or a new one on the next occasion [112,113]. It is an open question how this affects people’s sense of joint commitment in interactions with machines.

All this means that interactive interfaces provide a useful test bed for empirical work on how people construct meaning and coordinate action in interaction with artificial agents [114–116]. Such work will benefit from a conception of repair that focuses not just on how to deal with communicative trouble but also on how repair in human interaction helps organise and administer accountability.

Concluding remarks
Information theory has long acknowledged the importance of noise [117] and its role in shaping efficiency in communication systems [82]. A long-standing focus on individuals over interactions has led to the study of prevention mechanisms like efficient encoding [82,118] but has tended to overlook the equally consequential role of repair as an infrastructure for negotiating mutual understanding. Here, we show that a focus on interaction reveals how the linguistic system itself is shaped by and for its primary ecology. Languages adapt to the pressures of this ecology by evolving metacommunicative resources. Interactive repair allows people to communicate reliably without losing flexibility.

As cognitive science develops more interactive frameworks [119,120], repair will be central (see Outstanding questions). As we have shown, empirical and theoretical advances identify interactive repair as a domain of pragmatic universals in language use. Work on comparative cognition and human–computer interaction is providing a clearer view of features that are necessary and
sufficient for a general account of repair. Work on language development and joint action shows how repair can scaffold language learning and support collective computation. By augmenting and distributing cognitive processes, interactive repair effectively locates the ‘mystery’ of metacognition [121] in public interaction.

We conclude by emphasising two core properties of language that are criterial to interactive repair and that stand in a dialectical, mutually reinforcing relationship to each other. These are reflexivity and social accountability. Interactive repair both exploits and exhibits the unique reflexivity of language: its capacity to refer to and operate on itself. Interactive repair relies on and embodies methods for social accountability: it provides the techniques by which interactants can hold each other accountable and by which they can keep interactions on track. These ingredients of interactive repair are easily taken for granted, but they are indispensable in creating the power of interactive repair in human interaction, both to keep informational exchanges on track and to power the mutual accountability at the heart of our social relationships.

Acknowledgments
We are grateful to Marieke Woensdregt for feedback on parts of the manuscript, and to three anonymous reviewers for helpful comments. M.D. is funded by Dutch Research Council grant 016.wdi.185.205.

Declaration of interests
The authors have no interests to declare.

References