Uncovering the Rules of Entity-Level Robotic Working Memory

Rafael Sousa Silva (rsousasilva@mines.edu)
Tom Williams (twilliams@mines.edu)
MIRRORLab, Colorado School of Mines, 1500 Illinois Street
Golden, CO 80401 USA

Abstract

Working Memory (WM) is a necessary component for models of human cognition and human-inspired robot cognitive architectures. Different theories explain how the limited capacity of WM should be maintained, including theories of forgetting through decay and interference. Yet, it is unclear how WM models informed by these theories might be used to inform robot cognition, and how they might shape robots' ability to engage in natural, situated, language-based interactions. To resolve this tension, in this work we consider entity-level, feature-based WM systems that can be integrated into robot cognitive architectures to reflect both decay- and interference-based dynamics. We demonstrate how different parameterizations of these WM strategies have fundamentally different error modes in different interaction contexts. We formulate rules that inform the selection of decay and interference parameters to be used in contexts with different factors that are important for language-based interaction.

Keywords: robot cognitive architectures; working memory; decay; interference; referring expression generation

Introduction and Motivation

Robots are increasingly being deployed into situated task contexts in domains like education, healthcare, and space exploration (Gordon et al., 2016; Johanson et al., 2021; Roy et al., 2023), in which robots will need to communicate effectively with humans through natural language. A key enabler of human cognition is Working Memory (WM): a limited capacity storage cache and set of accompanying processes, which informs processes like reasoning (Kyllonen & Christal, 1990; Süß et al., 2002), comprehension (Halford et al., 1998), and learning (Baddeley, 2010), and linguistic processes like language generation (Gundel et al., 1993), understanding (Rönnberg et al., 2010), and acquisition (Baddeley et al., 1998; Denhovska et al., 2016). For robots to demonstrate these same capabilities, they too may need WM systems, whose design poses three key questions: (1) What is stored in WM? (2) What is the architecture of WM? (3) What are the dynamics of WM?

To answer the first key question, roboticists must determine what types of representations are stored in WM. While classic models of WM assumed that WM held a limited number of *entities*, more recent research has instead suggested that WM holds a limited number of *features* of those entities (Ma et al., 2014). While robots' WM systems do not necessarily need to directly match human WM systems in design, we argue that robotic WM systems should be *feature-based*, maintaining a *cache* of relevant features for relevant entities. This feature-based approach well aligns with the commitments of architectures like DIARC (Scheutz et al., 2019), where these features serve as the "common currency" for exchanging information between architectural components. Moreover, the feature-based approach may best allow WM to influence language generation, by influencing which features (rather than merely which entities) are referenced in robots' utterances.

To answer the second question, roboticists must determine how WM is coordinated across the robot architecture. Robots might use a *global* approach where a single pool of features is maintained in a centralized location. Second, robots might use a *component-level* approach where each robotic memory system (vision, mapping, etc.) maintains its own pool of features. Third, robots might use an *entity-level* approach where each robotic memory system maintains a distinct pool of features for each task-relevant entity. While this third option deviates from what is known about *human* cognition, it is a common approach for *robot* cognition as it leverages the unique properties of robotic architectures (Williams et al., 2018).

In this paper, we thus assume a feature-based, entity-level robotic WM system, and seek to answer the third question: *How should the dynamics of such a system be designed and parameterized?* As the defining feature of WM is its limited capacity, it is important to understand how long information should remain in WM and how WM contents affect natural-language-based robot cognitive processes. Thus, we focus our analysis on two key theories of *forgetting* from cognitive psychology (Reiter & Dale, 1997; Muter, 1980; Oberauer & Lewandowsky, 2014; Jonides et al., 2008; Lewandowsky et al., 2009). First, the theory of *decay* asserts that WM items are forgotten over time, if not rehearsed (Ebbinghaus, 1885; Brown, 1958). Second, the theory of *interference* asserts that the least recent unrehearsed information is removed when space is needed for new information (Waugh & Norman, 1965; Dewar et al., 2007).

To answer this third question, it is thus critical to understand how the values parameterizing decay- and interference-based robot cognitive models of WM should be selected. As we show in this work, the performance of these decay- and interference-based models are critically sensitive to the way those models are parameterized, and the relationship between those parameterizations and key dimensions of the contexts in which interaction unfolds.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

WM Systems in Robot Cognitive Architectures

To identify how the parameterizations of robotic models of WM impact natural language interactions, we must first clarify the set of assumptions we make about the fundamental nature of those WM models and how they might be implemented in robotics. In this section, we will thus explain the types of WM frameworks and WM forgetting models that can be used in robot cognitive architectures. Throughout this section, we will use the DIARC cognitive architecture (Scheutz et al., 2019) as an example to help clarify our ideas. DIARC implements key theories from cognitive psychology and linguistics to enable better language-capable robots, proving to be a good fit for the purposes of this paper.

Feature-Level Information

Within a robot cognitive architecture, WM knowledge can be represented with different levels of abstraction. A *feature-based* approach aims to prioritize the quality over the quantity of WM representations (Ma et al., 2014) by storing only the set of activated*features* that apply to entities (e.g., the predominant color of an entity, the entity's size and shape, etc). For the purposes of this paper, we consider only the features of entities that become activated through dialogue (cf. Higger & Williams (2024)). In DIARC, for example, information about entities is represented in first-order logic. In an interaction where a human asks a robot to hand them "the red mug on the counter," at a feature level the property `counter(Y)` becomes activated for the mentioned counter and the properties `red(X)`, `mug(X)`, and `on(X,Y)` become activated for the mentioned mug.

Entity-Level Dynamics

The distribution of WM buffers across a robot's architecture can also be implemented at different levels. For instance, some architectures may implement a single, *global* WM buffer of activated information. On the other hand, one WM buffer could be assigned to each different *entity type* or knowledge domain (e.g. people, locations, objects). In contrast, in an *entity-level* WM model each individual known entity has a dedicated WM buffer that stores activated content recently known to apply to that entity. We note that this is not necessarily a *cognitively plausible* account of WM, but it is a model that is well-tailored to robotics domains.

In DIARC, information about known entities is distributed across multiple knowledge bases rather than stored in a single database. This reduces the number of architectural bottlenecks and allows for information about different knowledge domains (e.g., objects, people, locations, etc) to be stored in different formats (Williams & Scheutz, 2016). At an entity level, each known entity inside the architecture will have an independent WM buffer storing activated content. If we consider again the interaction where a human asks a robot to hand them "the red mug on the counter," then in a knowledge base responsible for storing information about objects, the WM buffer for the counter will have `counter(Y)` in it while the buffer for the mug will separately store the features `red(X)`, `mug(X)`, and `on(X,Y)`.

Forgetting Strategies

Finally, while there are different ways in which temporal decay and interference can be designed in cognitive architectures, in this work we assume that *temporal decay* is implemented in a way such that the least recent (lowest activation) feature is removed from a WM buffer every δ seconds, ensuring the progressive removal of unrehearsed information from WM. Similarly, for *interference*, we assume that WM buffers can hold a maximum number of items, α, at any given time. When newly activated features need to be stored in WM, they replace the least recent (lowest activation) features in storage if space is needed.

DIARC's WM system is configured to reflect both decay- and interference-based dynamics. Its WM Manager component (Sousa Silva & Williams, 2024) manages a set of relevant WM buffers distributed across the entities known within the architecture. Features are added to an entity's WM buffer whenever that entity is mentioned in conversation, either by a human or by the robot itself, and features are removed from WM buffers according to the forgetting strategy in use.

DIARC's WM system is then used to facilitate tasks like Referring Expression Generation (REG) (cf. Van Deemter, 2016): when determining the features to use to refer to a target, those stored in WM are considered before querying Long Term Memory (LTM) for other features that might be used.

In this work we ask how a WM system's parameterization in terms of decay parameter δ or interference parameter α might be determined to best facilitate this REG process. To do so, we consider the distinct *error modes* that determine the success of REG under these forgetting strategies.

Decay Error Modes

We start by considering the error modes that would be associated with the decay forgetting strategy. That is, we must consider the distinct problems that can arise from using different δ values during an interaction. Since the implementation of WM decay that we consider is based on the intervals of time during which entity features remain in WM storage, we must identify the problems that can occur in interactive scenarios where information will remain in WM storage for too long, as well as scenarios in which information will leave WM too quickly. First, if δ is too high, then features will remain activated in WM storage for long time intervals, which can become a problem in dialogue if the features of the referred entities change. In these scenarios, robots may be prone to generating referring expressions containing outdated and invalid information. Second, if δ is too low, then the activated features of referred entities will not remain in WM for long enough to be useful, which can be problematic in interactions where a robot needs to refer to entities with sufficient frequency and leverage WM contents. As such, to choose an appropriate δ value for WM models that implement temporal decay, we need to take into account *how fast the environment* *is changing* and *how often an entity needs to be referred to* in conversation.

Environmental Dynamics

Let us first consider interactions in which the features of target referents dynamically change. When a robot is engaged in an interaction, the environment in which that interaction is taking place may change at different rates, and a feature that used to hold for an entity (or a relationship that used to hold between two entities) might not hold anymore. For example, the position of an object might change if someone moves it from its original location during dialogue. Consequently, robots may be prone to generating referring expressions containing outdated and invalid information *if* temporal decay is slower than the interval of environmental change, as the stale information will not leave buffers in time.

Naturally, the risk of using an outdated feature to describe an entity becomes higher as time goes on, suggesting it could be governed by an exponential process. For the purposes of this paper, we define this probability of using stale features, P_s, as a function of the decay time (δ) and the rate at which features are expected to go stale (λ_s), where higher $\lambda_s \in \mathbb{R}^+$ values will be seen in contexts where features go stale more quickly:

$$P_s(\delta;\lambda_s) = \begin{cases} 1 - e^{-\lambda_s \delta} & \text{if } \delta \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

This formulation allows us to specify the following rule for establishing an upper bound on δ:

Rule 1 *Given λ_s and a risk threshold $\bar{P}_s \in [0,1]$ specifying the maximum tolerable probability of a feature in WM going stale, we can then solve for an upper bound $\bar{\delta}$ such that $P_s(\delta;\lambda_s)$ is guaranteed to stay below $\bar{P}_s$, so long as $\delta \leq \bar{\delta}$.*

This relationship can be seen in Figure 1 for different values of λ_s.

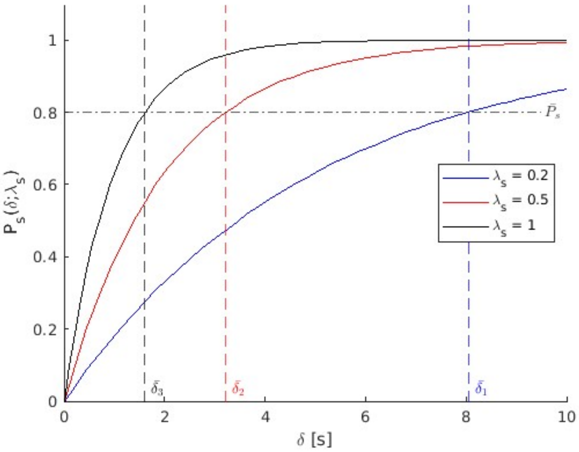

Figure 1: Environmental change example. If we let $\bar{P}_s = 0.8$, then the point where P_s and $\bar{P}_s$ intersect can be used to find $\bar{\delta}$.

Dialogue Dynamics

Remembering and reusing features that humans use to refer to specific entities can be helpful for language-capable robots, as their utterances may sound more natural, consistent, and familiar to their interlocutors. In addition, aggressive rates of decay may prevent the benefits of WM-facilitated REG. After all, $\delta = 0$ seconds would certainly avoid the use of stale features, but would do so by not keeping *anything* in WM. Thus, to promote the generation of natural and familiar referring expressions, we must find a lower bound on decay.

The risk of generating inconsistent referring expressions (i.e., referring expressions that fail to demonstrate entrainment to humans' lexical choices) decreases as decay time increases, suggesting it could be governed by a negative exponential process. In this paper, we define this probability of generating inconsistent referring expressions, P_r, as a function of decay time (δ) and the rate at which features are expected to be reused in dialogue (λ_r), where higher $\lambda_r \in \mathbb{R}^+$ values will be seen in contexts where features must be reused more quickly:

$$P_r(\delta;\lambda_r) = \begin{cases} e^{-\lambda_r \delta} & \text{if } \delta \geq 0 \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

Given this model, we can then find a boundary, $\underline{\delta}$, for the time that features should remain stored in an entity's WM buffer so that they can be reused. To do so, we specify the following rule for establishing a lower bound on δ:

Rule 2 *Given λ_r and a risk threshold $\bar{P}_r \in [0,1]$ specifying the maximum tolerable probability of generating inconsistent referring expressions, we can solve for a lower bound $\underline{\delta}$ such that $P_r(\delta;\lambda_r)$ is guaranteed to stay below $\bar{P}_r$, so long as $\underline{\delta} < \delta$.*

This formulation can be visualized in Figure 2 for different values of λ_r.

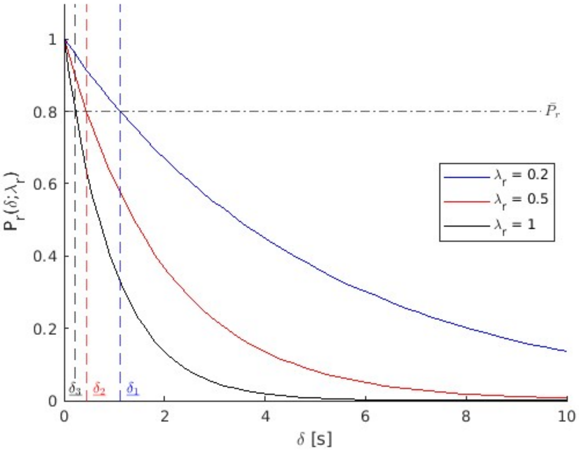

Figure 2: Dialogue dynamics example. If we let $\bar{P}_r = 0.8$, then the point where P_r and $\bar{P}_r$ intersect can be used to find $\underline{\delta}$.

Prioritization over Decay Rules

In certain situations when the chosen risk probability thresholds $\bar{P}_s$ and $\bar{P}_r$ are set to low values, the recommended upper bound $\bar{\delta}$ will be below $\underline{\delta}$, as shown in Figure 3. We argue that preventing the use of outdated, invalid features in dialogue is more important than preventing the use of inconsistent, yet valid features. As such, we specify the following metarule for decay:

Metarule 1 *In interaction contexts where $\underline{\delta} > \bar{\delta}$, an appropriate WM decay rate (δ) must be determined by satisfying **only** Rule 1. Otherwise, in interaction contexts where $\underline{\delta} \leq \bar{\delta}$, an appropriate δ must be determined by satisfying **both** Rule 1 and Rule 2.*

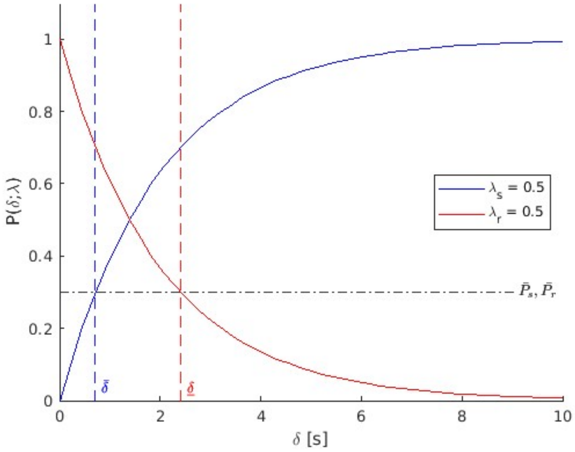

Figure 3: Example scenario where $\underline{\delta} > \bar{\delta}$. If we let $P_s = P_r = 0.3$ then no δ value can satisfy both Rule 1 and Rule 2.

Interference Error Modes

Let us now consider the error modes that would be associated with the interference forgetting strategy. To do so, we must identify the problems that can arise from situations where the maximum storage capacity of WM (α) is either too low or too high during interaction. On one hand, enough features must be stored in WM to facilitate the generation of unique referring expressions that discriminate the target referent from distractors and minimize the number of cache misses within WM. On the other hand, a higher WM buffer capacity increases the chances of storing stale features, and decreases the value of having an WM system to begin with, as WM comes to approximate the contents of LTM. Therefore, we argue that the appropriate choice of WM buffer size should depend on the minimum number or features typically needed to describe entities while ruling out all distractors in a given domain. In addition, we argue that the choice of maximum WM buffer size should account for the risk of stale features being stored in WM.

Minimizing Cache Misses

In Human-Robot Interaction (HRI) scenarios, when agents need to refer to specific entities they must minimize ambiguity. Ambiguous referring expressions can be problematic, as listeners may not be able to correctly identify the target referent, or may misinterpret speaker's utterances. We assume that when an entity e must be referred to, the contents of the WM buffer associated with e will be considered first during REG, and that LTM will only be consulted if those features stored in WM are insufficient to disambiguate e from its distractors (Williams et al., 2018; Sousa Silva et al., 2023). Therefore, if we define a boundary for the minimum WM buffer capacity (α) that depends on the minimum number of features typically needed to uniquely describe entities, we can minimize the number of LTM queries that are needed overall.

This minimum number of features needed for REG ultimately depends on the complexity of the domain to which target referents belong. First, specific types of attributes may be used more often than others to describe different types of entities. For example, color and shape may be used more frequently to describe objects than to describe people or locations. Second, the number of distractors in the domain affects the number of features that must be used to uniquely describe a target referent. Given this dependence on domain complexity, we argue that the lower bound for WM buffer capacity ($\underline{\alpha}$) should be tailored to each specific domain through data-driven estimation of how many descriptors, on average, are needed to describe an entity in that domain.

In previous work, for example, Piwek (2007) characterized the number of features used in observed referring expressions within the context of a common reference game in which one speaker delivers monologues to enable a listener to construct a desired shape from colored blocks (cp. Han et al., 2022). The authors observe that in their domain, the average number of features included in purely verbal referring expressions was 1.7, and that this number of features followed a distribution in which 90% of purely verbal referring expressions involved only one or two features. With this example in mind, we can formalize our first interference rule:

Rule 3 *Given the average number of features needed to uniquely describe entities in a given domain ($\hat{\alpha}$), we can define a lower bound for WM buffer capacity as $\underline{\alpha} = \lceil \hat{\alpha} \rceil$.*

Environmental Dynamics

Finally, just as high decay time values present risk of stale feature use, P_s, so too do high WM buffer capacity limits. With higher WM buffer capacities, not only does the risk of holding stale features in WM increase, but also WM becomes closer to functioning as a second LTM storage. We must then find a boundary, $\bar{\alpha}$, for the maximum WM buffer capacity without risking an unacceptable chance of storing stale descriptors.

We can then define this interference-based probability of storing stale features in WM, P_i, as a function of the average amount of time that features are expected to remain in WM storage ($\hat{\delta}$) and the rate at which features are expected to go stale (λ_s).

$$P_i(\hat{\delta};\lambda_s) = \begin{cases} 1 - e^{-\lambda_s \hat{\delta}} & \text{if } 0 \leq \alpha \leq N \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Here, $\hat{\delta}$ is calculated as $\hat{\delta} = \frac{\alpha}{\lambda_f}$, where α is the WM buffer capacity for each entity, λ_f is the rate at which individual features are expected to leave WM, and N is the maximum number of features that might ever need to be used, as described below.

λ_f may in turn be calculated as $\lambda_f = \lambda_r P_f$, where λ_r is the rate at which features are expected to be reused in dialogue, and P_f is the probability that a feature will be removed from WM storage to make room for a new feature upon each reference.

Finally, P_f is calculated by assuming that all types of features are equally likely to be used, and that a feature is removed from WM when a new feature of a type not already in WM storage needs to be inserted yet no free WM slots are available.

Calculating the probability that a feature *of a type not already appearing in WM* will be used requires knowledge of the total number of feature types N that might be used. While hypothetically in highly complex environments a large number of features and spatial relations could in theory be needed, it may also be reasonable to assume a tractable value of N. For example, a value such as seven could be justified by analysis of the ReferIt dataset, which suggests that when describing visible objects in natural scenes, humans essentially use only seven key types of attributes (Kazemzadeh et al., 2014).

Given N, P_f can be calculated as

$$1 - \sum_{n=1}^{\alpha} P(R_e = WM_e[n]),$$

That is, the probability that, upon feature R_e being used to refer to entity e, none of the features in entity e's WM buffer are of the same type as R_e. Under the assumption of uniform feature use, $P(R_e = WM_e[n]) = \frac{1}{N}$. This equation can be approximated as $P_f = 1 - \frac{\alpha}{N}$.

Therefore, the exponent from Equation 3 can be simplified as follows:

$$-\lambda_s\hat{\delta} = \frac{-\lambda_s\alpha}{\lambda_f} = \frac{-\lambda_s\alpha}{\lambda_r P_f} = \frac{-\lambda_s\alpha}{\lambda_r(1-\frac{\alpha}{N})} = (\frac{-\lambda_s}{\lambda_r})(\frac{\alpha N}{N-\alpha})$$

We can thus define the interference-based probability of using stale features, P_i, as a function of the WM buffer capacity (α), the rate at which features are expected to go stale (λ_s), the rate at which features are expected to be reused in dialogue (λ_r), and N:

$$P_i(\alpha;\lambda_s;\lambda_r;N) = \begin{cases} 1 - e^{(\frac{-\lambda_s}{\lambda_r})(\frac{\alpha N}{N-\alpha})} & \text{if } 0 \leq \alpha \leq N \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

We thus arrive at our final interference rule:

Rule 4 *Given λ_s, λ_r, N, and a risk threshold $\bar{P_i} \in [0,1]$ specifying the maximum tolerable probability of a feature in WM going stale, we can solve for an upper bound $\bar{\alpha}$ such that $P_i(\alpha;\lambda_s;\lambda_r;N)$ is guaranteed to stay below $\bar{P_i}$, so long as $\alpha \leq \lceil\bar{\alpha}\rceil$.*

This relationship can be visualized in Figure 4 for different values of λ_s.

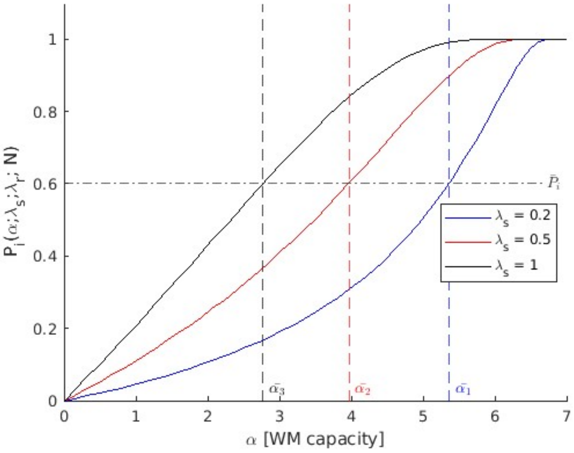

Figure 4: Interference environmental dynamics example. If we let $\bar{P_i} = 0.6$, $\lambda_r = 0.2$, and $N = 7$, then the point where $\bar{P_i}$ and P_i intersect can be used to find $\bar{\alpha}$.

Prioritization over Interference Rules

Similarly to decay, in scenarios with a lower risk probability threshold, $\bar{P_i}$, our rule system fails to provide a valid range for appropriate α values. Thus, we must adjust our interference rules to real-world scenarios with fast environmental change and consider whether minimizing ambiguity and cache misses outweighs preventing the use of stale features in robot referring expressions further. While the former risk is still associated with the formulation of valid referring expressions, the latter risk is not. This reinforces the idea that robots must prioritize avoiding the use of stale features in dialogue, as their utterances may lose their validity. Therefore, our interference metarule also ensures that robot dialogue will prioritize avoiding the use stale features:

Metarule 2 *In interaction contexts where $\underline{\alpha} > \bar{\alpha}$, an appropriate WM buffer size (α) must be determined by satisfying **only** Rule 4. Otherwise, in interaction contexts where $\underline{\alpha} \leq \bar{\alpha}$, an appropriate α must be determined by satisfying **both** Rule 3 and Rule 4.*

Discussion

Contributions — The main contribution of this paper is the introduction of metarules that bound the parameters that guide decay and interference in entity-level, feature-based resource management strategies for the WM systems of integrated robot architectures. The values of these bounds are based on key identified features of interaction contexts. Our work provides roboticists with clear and justified guidance for setting these parameters to minimize sources of risk arising from the use of WM systems.

Furthermore, while the WM systems described in this work are tailored to robot cognitive architectures and are not cognitively plausible, the metarules presented in this paper are nevertheless of interest from a broader cognitive science perspective, as our formulation of these metarules demonstrates the relationship between the parameterization of decay and interference based WM models and key facets of the domains in which human interaction unfolds. Although the exact parameterization of cognitively plausible WM models would undoubtedly differ, a similar relationship may hold between decay and interference levels in more cognitively plausible models and these dimensions of situated cognition and interaction. Our metarules thus suggest the need to investigate similar parameterization questions in the context of computational cognitive models of human cognition.

Limitations and Future Work — One key limitation of our approach is the assumption of stable, context-specific rates of environmental and dialogue dynamics. While it may be reasonable for these rates to be specified or learned at a context level, these rates might well vary for different types of objects within a single context. In an assembly task, for instance, it might be beneficial for a robot to keep salient features of *task-relevant* objects in WM for longer than other, incidental features of the location where the task is being performed. More natural dialogue might thus be achieved through a dynamic decay policy that can be adjusted to handle different types of objects with different decay rates. Alternatively, these rates might be learned or specified at an even deeper level, by specifying or learning different rates of expected change and reference for different types of *object features*. A robot could, for example, learn to maintain object type in WM for a longer duration, while allowing object position to decay more rapidly. Future work can thus investigate how different decay rates might be specified or learned at the level of object features.

Another key limitation of this work is the lack of empirical validation. While the rule derivations above formally demonstrate the relationships between dimensions of situated contexts, parameter settings, and the risks of resulting error modes, future work is needed to empirically validate this model, for two key purposes. First, models parameterized in ways that adhere to or violate our proposed rules should be evaluated with human subjects to determine the accuracy of predicted rates of encountering error modes, and to determine the extent to which these error modes lead to observable problems for human participants. Second, while in this work we are not focused on cognitively plausible models of human cognition, assessment of the fit of the proposed models to human data would provide valuable insights both for roboticists and cognitive scientists.

Conclusion

In this paper, we derived parameter selection rules for an *Entity-Level, Feature-Based* WM framework for robotic cognitive architectures applied to natural, situated, language-based interaction scenarios with humans. Our rules are designed around cognitively-inspired implementations of decay and interference, which are forgetting dynamics that dictate how information leaves WM buffers. In addition, our rules consider different error modes for decay and interference that can affect a robot's Referring Expression Generation process. That is, we identified situations in which using specific decay and interference parameterizations might be problematic in HRI contexts. We derived metarules based on these parameters to promote the generation of robot referring expressions that will sound intuitive, natural, and easy to understand for human interactants. We hope our rules can inform future human-subjects experiments aimed at assessing how humans will perceive the referring expressions that are generated through this framework.

Acknowledgments

This work was funded in part by NSF CAREER grant IIS-2044865. Furthermore, we would like to extend our deepest thanks to the reviewers of this paper, who provided suggestions for major revisions to our manuscript, which resulted in a substantially improved paper relative to our original version.

References

Baddeley, A. (2010). Working memory. *Current biology*, *20*(4).

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Exploring Working Memory*, 164–198.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly journal of experimental psychology*, *10*(1), 12–21.

Denhovska, N., Serratrice, L., & Payne, J. (2016). Acquisition of second language grammar under incidental learning conditions: The role of frequency and working memory. *Language Learning*, *66*(1), 159–190.

Dewar, M., Cowan, N., & Della Sala, S. (2007). Forgetting due to retroactive interference: A fusion of müller and pilzecker's (1900) early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex*, *43*(5), 616–634.

Ebbinghaus, H. (1885). Memory: A contribution to experimental psychology, trans. *HA Ruger & CE Bussenius. Teachers College.*.

Gordon, G., Spaulding, S., Westlund, J., Lee, J., Plummer, L., Martinez, M., … Breazeal, C. (2016). Affective personalization of a social robot tutor for children's second language skills. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 30).

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274–307.

Halford, G., Wilson, W., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and brain sciences*, *21*(6), 803–831.

Han, Z., Rygina, P., & Williams, T. (2022). Evaluating referring form selection models in partially-known environments. In *Proceedings of the 15th international conference on natural language generation* (pp. 1–14).

Higger, M., & Williams, T. (2024). GAIA: A Givenness Hierarchy Theoretic Model of Situated Referring Expression Generation. In *Proceedings of the annual meeting of the cognitive science society.*

Johanson, D., Ahn, H., & Broadbent, E. (2021). Improving interactions with healthcare robots: a review of communication behaviours in social and healthcare contexts. *International Journal of Social Robotics*, *13*(8), 1835–1850.

Jonides, J., Lewis, R., Nee, D., Lustig, C., Berman, M., & Moore, K. (2008). The mind and brain of short-term memory. *Annual review of psychology*, *59*, 193.

Kazemzadeh, S., Ordonez, V., Matten, M., & Berg, T. (2014). Referitgame: Referring to objects in photographs of natural scenes. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 787–798).

Kyllonen, P., & Christal, R. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433.

Lewandowsky, S., Oberauer, K., & Brown, G. (2009). No temporal decay in verbal short-term memory. *Trends in cognitive sciences*, *13*(3), 120–126.

Ma, W. J., Husain, M., & Bays, P. (2014). Changing concepts of working memory. *Nature neuroscience*, *17*(3), 347–356.

Muter, P. (1980). Very rapid forgetting. *Memory & Cognition*, *8*(2), 174–179.

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language*, *73*, 15–30.

Piwek, P. (2007). Modality choice for generation of referring acts: Pointing versus describing.

Reiter, E., & Dale, R. (1997). Building applied natural language generation systems. *Natural Language Engineering*, *3*(1), 57–87.

Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*, *12*(49), 263.

Roy, S., Smith, T., Coltin, B., & Williams, T. (2023). I Need Your Help... or Do I? Maintaining Situation Awareness Through Performative Autonomy. In *Proceedings of the 2023 acm/ieee international conference on human-robot interaction* (pp. 122–131).

Scheutz, M., Williams, T., Krause, E., Oosterveld, B., Sarathy, V., & Frasca, T. (2019). An overview of the distributed integrated cognition affect and reflection diarc architecture. *Cognitive architectures*, 165–193.

Sousa Silva, R., Lieng, M., & Williams, T. (2023). Forget about it: Entity-level working memory models for referring expression generation in robot cognitive architectures. In *Proceedings of the annual meeting of the cognitive science society.*

Sousa Silva, R., & Williams, T. (2024). Say what? analyzing the impact of an entity-level model of working memory forgetting on referring expression generation.

Süß, H.-M., Oberauer, K., Wittmann, W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*(3), 261–288.

Van Deemter, K. (2016). *Computational models of referring: a study in cognitive science*. MIT Press.

Waugh, N., & Norman, D. (1965). Primary memory. *Psychological review*, *72*(2), 89.

Williams, T., & Scheutz, M. (2016). A framework for resolving open-world referential expressions in distributed heterogeneous knowledge bases. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 30).

Williams, T., Thielstrom, R., Krause, E., Oosterveld, B., & Scheutz, M. (2018). Augmenting robot knowledge consultants with distributed short term memory. In *International conference on social robotics* (pp. 170–180).