Competence-Aware Autonomy: An Essential Skill for Robots in the Real World

Connor Basich,[1] Joydeep Biswas,[2] and Shlomo Zilberstein[1]

[1] University of Massachusetts Amherst
[2] The University of Texas at Austin
cbasich@cs.umass.edu, joydeepb@cs.utexas.edu, shlomo@cs.umass.edu

Abstract

Recent efforts in AI and robotics towards deploying intelligent robotic systems in the real world offer the possibility of transformational impacts on society. For such systems to be successful while reliably maintaining safe operation, they must be cognizant of their limitations, and when uncertain about their autonomous capabilities, solicit human assistance. However, system designers cannot fully enumerate the space of all situations that a robot deployed in the real world might face, prompting the challenge of endowing robots with actionable awareness of their capabilities and limitations in unseen settings. We propose *competence-aware autonomy* as a means of addressing this challenge in a well-defined manner motivated by real world examples. We discuss recent prior work in this area and suggest several research challenges and opportunities for future work.

1 Introduction

The recent growth in artificial intelligence and robotics has spurred new, increasingly ambitious efforts towards deploying intelligent robotic systems in complex real world environments. These systems can assist humans across a wide spectrum of tasks, offering the possibility of transformative changes that have the potential to alter our society and our everyday lives. Examples of such efforts include unmanned underwater vehicles (Kunz et al. 2009), extraterrestrial space rovers (Gao and Chien 2017), service robots (Biswas and Veloso 2016; Hawes et al. 2017), and self-driving cars (Badue et al. 2021).

However, the open world poses new challenges such as unconstrained and non-stationary environments, multiple heterogeneous actors, partial observability, and unexpected scenarios. Individually, each challenge may influence the way in which we frame the problem, but the combination of multiple such challenges calls into question many of the traditional assumptions used in AI and robotics. This is compounded by the fact that many of these systems are expected to operate *safely* and *reliably* on the order of months or even years. Consequently, in order to execute their tasks safely and reliably, such robots are often deployed with some capacity to rely on various forms of human assistance that can support the robot in completing its task. However, even with the support of expert humans, robotic systems deployed for long periods in challenging open-world domains are likely to encounter unanticipated situations that were not explicitly accounted for *a priori* by the robot developers.

How can a robot predict that its perceptual estimates are in error, if it is not pre-programmed with models of such errors? How can it identify the root causes of such failures, so that when encountering another error it could reason about whether the two are instances of the same type of scenario? Furthermore, if it is not programmed to even detect such errors, how could it ask for help from humans to overcome them? And finally, how can it determine the best type of help to ask for, if the situations were not considered *a priori*?

To handle these challenges, we propose *competence-aware autonomy*, enabling intelligent robots to learn and reason about (1) their limitations in executing a task autonomously, (2) the environmental or situational factors that influence these limitations, and (3) the proper form and extent of human assistance to request to optimally compensate for their limitations. We argue that an intelligent robot deployed into the real world should be able to reason, at any point, about whether it has the requisite *competence* to act autonomously, and, if not, reason about the appropriate level of human assistance needed to compensate for its limited competence. In particular, the robot should aim to not be over-reliant on human assistance, placing unnecessary burden on the human that may lead to a higher cost, an overburdened human, and potentially lower trust in the system and less willingness to use the system. At the same time, the robot should also aim to not be under-reliant on human assistance, taking excessive time and energy to perform what may be a simple, low-cost operation for the human, or worse, attempting what may be an unsafe operation for the robot.

The rest of this paper is structured as follows: in Section 2 we define competence-aware autonomy; in Section 3 we define competence-aware perception and discuss relevant work in the area; in Section 4 we define competence-aware planning and discuss relevant work in the area; in Section 5 we introduce future challenges in the area and some ideas on how they may be approached; finally, in Section 6 we offer concluding thoughts on the content introduced in the paper.

2 Competence-Aware Autonomy

Motivated by earlier studies of *competence* in the context of human workers (Further Education Unit 1984; Sternberg and Kolligian Jr. 1990; Hager and Gonczi 1996; Gilbert

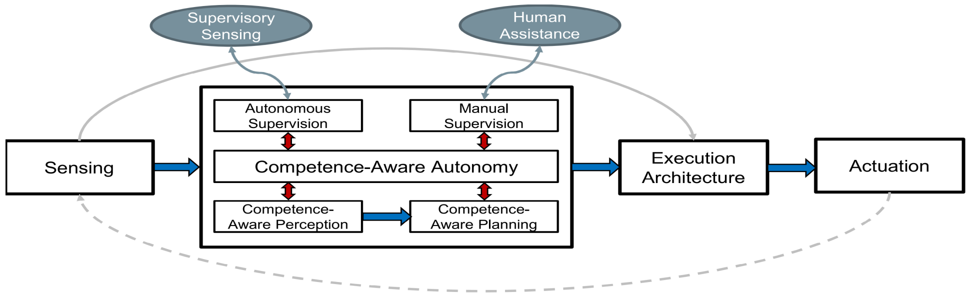

Figure 1: Competence-Aware Autonomy extends the traditional Sense, Plan, Act framework.

2013), we propose that competence-aware autonomy (Figure 1) is, intuitively, the ability of an agent to know, reason about, and, ultimately, act on the extent of its own capabilities in any situation in the context of different sources of external assistance. This is an ability that we as humans frequently utilize. In the course of our everyday lives we are constantly reasoning over what we know and what we can do when attempting to complete a task, and when those abilities fall short of ensuring that the task can be done in a safe and successful manner, we must then reason about what assistance we can seek. However, even for humans this knowledge is imperfect and must be learned over time through experience and feedback from others, and updated as our own abilities change. Similarly a robot must have the ability to learn and reason about its own competence. For a robot to achieve competence-aware autonomy, it must have the capability to reason about both its *perceptual competence* and *planning competence*.

Perceptual competence is the ability of a system to infer both the states of the world accurately, as well as provide well-calibrated estimates of uncertainty as a function of sensed data. A system that has the capacity to introspectively reason about its perceptual competence to both determine and predict whether it has perceptual competence in some task is said to have *competence-aware perception*.

Planning competence is the ability of a system to determine—and proactively account for—whether autonomous operation is likely to succeed (in a safe and reliable manner) when executing an action, and when not, to reason over what form of human assistance may be needed, and how to obtain that human assistance in a timely manner. A system that has the capacity to introspectively reason about its planning competence when planning in the context of different forms of human assistance is similarly said to exhibit *competence-aware planning*.

Failure to account for one or both of these sources of competence can reduce the applicability of these robot's to real-world domains where the robots *a priori* models are imperfect but the system must reliably exhibit safe behavior.

For example, consider the scenario depicted in Figure 2 of an autonomous vehicle (AV) approaching a stop sign at an intersection (a). The AV believes that there is no oncoming traffic (as it detects none), but recognizes that its view to the right is partially obstructed (b). Competence-aware perception should reason about the existence of the foliage, and, more importantly, that the foliage may call into question the validity of its object sensing results. The first time that this scenario is encountered, the AV should solicit help from a human operator (either in the vehicle, or remote) (c) who can instruct the AV with a different action (edge forward) to resolve the potential error source (d). Finally, after encountering the scenario several times, the AV should recognize both the situation *and* how to resolve it, and competence-aware planning should proactively propose the potential resolution to the human instead of the potential error (e) to acquire a swift approval/disapproval response (f).

3 Competence-Aware Perception

Robots deployed in uncontrolled real-world settings such as commercial establishments and urban environments will inevitably encounter scenarios that violate the assumptions of their deployers, leading to execution failures. A common source of such failures is *perceptual errors*—where a perception algorithm provides estimates that are inaccurate, or inconsistent with the real world. While there exist previous works on uncertainty quantification, they rely on model-based uncertainty such as the Cramer-Rao Lower Bound for simultaneous localization and mapping (SLAM) (Pandey et al. 2015), hand-crafted measures of uncertainty (Sadat et al. 2014; Mostegel, Wendel, and Bischof 2014), or computationally expensive estimates for neural-network based perception such as Bayesian Neural Networks (BNNs) (Ghahramani 1997; Dusenberry et al. 2020) and Monte-Carlo Dropouts (Gal and Ghahramani 2016).

Figure 2: Competence-aware prediction of failures and integration of learned model into planning and execution.

To tackle the problem of identifying and overcoming causes of perceptual errors in novel environments, we leverage *introspective perception* (Daftry et al. 2016; Rabiee and Biswas 2019; Rabiee and Biswas 2020; Rabiee et al. 2022) as a general formulation for robots to autonomously identify causes of failures using either consistency across sensing modalities, or spatio-temporal consistency. We present three specific ways in which introspective perception may be used to overcome errors in obstacle avoidance, localization and mapping, and navigation.

Introspective Vision for Obstacle Avoidance

Vision-based obstacle detection algorithms rely on algorithmic assumptions and simplifications for computational tractability—for example that surfaces are lambertian and texture-rich, or that there are no aliasing features or refractive surfaces. When such assumptions are violated, the perception algorithms produce erroneous estimates, either hallucinating obstacles that do not exist (false positives), or missing obstacles that do exist (false negatives).

Introspective Vision for Obstacle Avoidance (IVOA) (Rabiee and Biswas 2019) overcomes such limitations by leveraging a supervisory sensor that is occasionally available. By comparing plans generated using the supervisory sensor and the plans generated using the vision algorithm under test, IVOA is able to identify scenarios where the vision algorithm produces erroneous results. IVOA projects the 3D coordinates where the plans agree and disagree onto the images used for vision-based perception to generate a training set of reliable and unreliable image patches for perception. IVOA then uses this training dataset to learn a model of which image patches are likely to cause failures of the vision-based obstacle detection. Using this model, IVOA is able to predict whether the relevant image patches in the observed images are likely to cause failures.

Introspective Vision for Simultaneous Localization and Mapping

Visual simultaneous localization and mapping (V-SLAM) algorithms most commonly assume that errors in feature extraction and matching are independent and identically distributed (i.i.d), but this assumption is often violated (Triggs et al. 1999)—for example, features extracted from low-contrast regions of images exhibit wider error distributions than features from sharp corners. Furthermore, V-SLAM algorithms are prone to catastrophic tracking failures when sensed images include challenging conditions such as specular reflections, lens flare, or shadows of dynamic objects. Previous work to address these failures has focused on building more robust visual frontends.

Introspective Vision for SLAM (IV-SLAM) is a fundamentally different approach to these challenges. IV-SLAM explicitly models the noise process of re-projection errors from visual features to be context-dependent, and hence non-i.i.d. IV-SLAM leverages spatio-temporal consistency as an autonomously supervised approach to collect training data to learn such a context-aware noise model. Using this learned noise model, IV-SLAM guides feature extraction to select more features from parts of the image that are likely to result in lower noise, and further incorporate the learned noise model into the joint maximum likelihood estimation, thus making it robust to the aforementioned types of errors.

Competence-Aware Planning via Introspective Perception

While introspective vision of obstacle avoidance is effective at reasoning about failures of perception for obstacle detection, it does not necessarily directly translate to reasoning about failures of *task execution*. For example, while a robot may have errors in accurately estimating whether a lane has a parked car or not, if the navigation task does not require the robot to enter the lane, the perceptual uncertainty would not impact the navigation task success. Competence-Aware Planning via Introspective Perception (CPIP) (Rabiee et al. 2022) instead explicitly reasons about the relation between perceptual failures and task-level competence by factorizing the competence-aware planning problem into two components. First, perception errors are learned in a model-free and location-agnostic setting via *introspective perception* prior to deployment in novel environments. Second, during actual deployments, the prediction of task-level failures is learned in a context-aware setting as the transition function in a stochastic shortest path problem that captures the probability of different classes of failures for each action.

4 Competence-Aware Planning

Intuitively, *competence-aware planning* is the introspective ability to optimally manage the reliance on human assistance. Reliance on human assistance in AI has been thoroughly studied over the years, and has often been predicated on the notion of *levels of autonomy* which is a paradigm for modeling a discretized representation of different limitations on autonomous operation, and their commensurate level of human assistance within a human-agent team (Parasuraman, Sheridan, and Wickens 2000; SAE On-Road Automated Vehicle Standards Committee 2014; Beal and Rogers 2020).

Mixed-initiative control (Allen, Guinn, and Horvtz 1999; Ghalamzan et al. 2017) considers human-agent systems in which each actor may take the initiative to act at different stages of control to best utilize their respective abilities, and has recently been applied in the context of variable autonomy in which the level of autonomy can change dynamically (Chiou, Hawes, and Stolkin 2021). Symbiotic autonomy (Rosenthal, Biswas, and Veloso 2010; Veloso et al. 2015) aims to design human-agent systems in which the human and agent act asynchronously to achieve their own goals but may perform tasks for each other to complete their tasks more efficiently overall. Adjustable autonomy (or, sometimes, variable autonomy) is a closely related area of research that has been extensively studied over the years (Mostafa, Ahmad, and Mustapha 2019) and generally considers human-agent teams that are specifically characterized by the ability to dynamically change between different levels, or modes, of autonomy each of which corresponds to different constraints on the system that affect the actions that the human-agent team can perform.

Competence-Aware Systems

A *competence-aware system* (Basich et al. 2020) is a competence-aware planning framework predicated on the ability of an agent to operate in different levels of autonomy in which the agent can proactively reason about the true, optimal extent of autonomy that the agent can and should utilize in any given situation. The *competence* of a CAS is specifically defined to be the optimal (i.e. reward-maximizing or cost-minimizing) level of autonomy to perform a given action in a given state, were the CAS to have a perfect model of the human's feedback and assistance. A competence-aware system learns its planning competence through interactions with a human which provide *feedback* in the form of discretized signals. By incorporating a predictive model of these feedback signals, a CAS can make an informed decision during planning about which level to employ in each situation, given its current constraints on which levels of autonomy it is allowed to use.

Given sufficient feedback from the human, a competence-aware system can be show to converge to its competence, meaning that it has learned to optimize its reliance on human assistance, resulting in the most efficient and cost-effective performance when completing its tasks. However, the learned competence itself, a function of the true predictive model of human feedback and assistance, may be poor when the system's model does not align well with the model of the human operator who provides the feedback from which the CAS learns. This phenomenon can arise in practice as it is unlikely that the human's model will perfectly align with the system's, and particularly when the human agent in the system is not a designer of the system. The result is that the feedback can appear random or inconsistent, leading to low competence and inefficient operation.

To address this, recent work introduced a method called *iterative state space refinement* (Basich et al. 2021) that enables the competence-aware system to refine the granularity of its state representation through online model updates by identifying cases where the system's model may be missing important features and integrating them into its model. This method leads to more fine-grained partitioning of the input space that can learn a more nuanced competence model, leading to improved performance and a better trade-off between autonomy and human assistance.

5 Future Challenges

Thus far, we have treated competence-aware perception and competence-aware planning as isolated components of an overall system's competence-aware autonomy. However, it is clear to see that a system's perceptual competence is a fundamental component of its planning competence, as it drives the system's capacity to autonomously operate safely and reliably in different situations in the context of potential perception-driven errors. Similarly, competence-aware planning should reason about how to effectively manage limited perceptual competence, providing corrective signals in the form of actions, options, or full contingency plans, to help either refine the competence-aware perception model, or identify new causal factors.

We propose that an important challenge moving forward is to endow such systems with an effective means of "closing the loop" between competence-aware perception and competence-aware planning, such that the system can rely on them to work synergistically to overcome errors in new environments. To this end, we identify two directions of interaction between the two competence-aware components.

Perception Abstractions for Planning We consider a hierarchical planning framework that performs planning at multiple levels of abstraction in parallel: mission-level, task-level, and situation-level. Competence-aware perception provides abstracted state information used at the situation-level (of which there may be multiple) which should be masked out for the specific action being considered by the given situation-level planner. However, higher levels of planning abstraction (i.e. task and mission) should *incorporate* perceptual competence into their planning models to ensure that global planning is robust to situations with low perceptual competence, and optimizes the policies generated with respect to the known perceptual competence model.

Repair-Policies Once competence-aware planning has learned to identify common scenarios of low competence, as presented in the example from Figure 2, the Repair-Policy pattern proceeds in five phases: (1) the situation-level planner encounters a situation where perception provides abstracted observations with low reliability predicted by the competence model; (2) competence-aware perception further identifies a causal factor (or factors) for the low reliability; (3) the situation-level planner solicits assistance from the human operator; (4) over repeated instances where the same causal factor is identified as responsible given the same abstracted observation in the same or similar part of the state space, and when the resultant human feedback is the same, competence-aware planning should hypothesize a *repair policy*; and (5) the repair policy should be presented to the human operator and, if approved, be used by the robot in the future when encountering the same scenario.

6 Conclusion

In this position paper, we outline some key challenges faced by robotic systems deployed in complex open-world environments, particularly challenges arising from being presented by a novel and unanticipated scenario. To address these challenges, we propose *competence-aware autonomy*, the ability of an autonomous agent to introspectively learn, model, and reason about its own capabilities in the context of different sources of external assistance. We present recent work on competence-aware autonomy, specifically in the areas of *competence-aware perception* and *competence-aware planning*, and discuss some of the key challenges moving forward in making competence-aware autonomy more robust to real-world challenges by closing the loop on both elements of competence-awareness.

7 Acknowledgments

This work was supported by the National Science Foundation grant IIS-1954782.

References

Allen, J. E.; Guinn, C. I.; and Horvtz, E. 1999. Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications*, 14(5): 14–23.

Badue, C.; Guidolini, R.; Carneiro, R. V.; Azevedo, P.; et al. 2021. Self-driving cars: A survey. *Expert Systems with Applications*, 165: 113816.

Basich, C.; Svegliato, J.; Beach, A.; Wray, K. H.; Witwicki, S.; and Zilberstein, S. 2021. Improving competence via iterative state space refinement. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1865–1871.

Basich, C.; Svegliato, J.; Wray, K. H.; Witwicki, S.; Biswas, J.; and Zilberstein, S. 2020. Learning to optimize autonomy in competence-aware systems. In *International Conference on Autonomous Agents and MultiAgent Systems*, 123–131.

Beal, J.; and Rogers, M. 2020. Levels of autonomy in synthetic biology engineering. *Molecular Systems Biology*, 16(12): e10019.

Biswas, J.; and Veloso, M. 2016. The 1,000-km Challenge: Insights and Quantitative and Qualitative Results. *IEEE Intelligent Systems*, 31(3): 86–96.

Chiou, M.; Hawes, N.; and Stolkin, R. 2021. Mixed-initiative variable autonomy for remotely operated mobile robots. *ACM Transactions on Human-Robot Interaction*, 10(4): 1–34.

Daftry, S.; Zeng, S.; Bagnell, J. A.; and Hebert, M. 2016. Introspective perception: Learning to predict failures in vision systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1743–1750.

Dusenberry, M.; Jerfel, G.; Wen, Y.; Ma, Y.; Snoek, J.; Heller, K.; Lakshminarayanan, B.; and Tran, D. 2020. Efficient and scalable Bayesian NN with rank-1 factors. In *International Conference on Machine Learning*, 2782–2792.

Further Education Unit. 1984. Towards a Competence-Based System: An FEU View.

Gal, Y.; and Ghahramani, Z. 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*, 1050–1059.

Gao, Y.; and Chien, S. 2017. Review on space robotics: Toward top-level science through space exploration. *Science Robotics*, 2(7).

Ghahramani, Z. 1997. Learning dynamic Bayesian nets. In *International School on Neural Networks*, 168–197.

Ghalamzan, E. A. M.; Abi-Farraj, F.; Giordano, P. R.; and Stolkin, R. 2017. Human-in-the-loop optimisation: Mixed initiative grasping for optimally facilitating post-grasp manipulative actions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3386–3393.

Gilbert, T. F. 2013. *Human competence: Engineering worthy performance*. John Wiley & Sons.

Hager, P.; and Gonczi, A. 1996. What is competence? *Medical Teacher*, 18(1): 15–18.

Hawes, N.; Burbridge, C.; Jovan, F.; Kunze, L.; Lacerda, B.; Mudrova, L.; Young, J.; Wyatt, J.; Hebesberger, D.; Kortner, T.; et al. 2017. The STRANDS project: Long-term autonomy in everyday environments. *IEEE Robotics & Automation Magazine*, 24(3): 146–156.

Kunz, C.; Murphy, C.; Singh, H.; Pontbriand, C.; Sohn, R. A.; Singh, S.; Sato, T.; Roman, C.; Nakamura, K.; Jakuba, M.; et al. 2009. Toward extraplanetary under-ice exploration: Robotic steps in the Arctic. *Journal of Field Robotics*, 26(4): 411–429.

Mostafa, S. A.; Ahmad, M. S.; and Mustapha, A. 2019. Adjustable autonomy: A systematic literature review. *Artificial Intelligence Review*, 51(2): 149–186.

Mostegel, C.; Wendel, A.; and Bischof, H. 2014. Active monocular localization: Towards autonomous monocular exploration for multirotor MAVs. In *IEEE International Conference on Robotics and Automation*, 3848–3855.

Pandey, G.; McBride, J. R.; Savarese, S.; and Eustice, R. M. 2015. Automatic extrinsic calibration of vision and LiDAR by maximizing mutual information. *Journal of Field Robotics*, 32(5): 696–722.

Parasuraman, R.; Sheridan, T. B.; and Wickens, C. D. 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30(3): 286–297.

Rabiee, S.; Basich, C.; Wray, K. H.; Zilberstein, S.; and Biswas, J. 2022. Competence-aware path planning via introspective perception. *IEEE Robotics and Automation Letters*, 7(2): 3218–3225.

Rabiee, S.; and Biswas, J. 2019. IVOA: Introspective vision for obstacle avoidance. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1230–1235.

Rabiee, S.; and Biswas, J. 2020. IV-SLAM: Introspective vision for simultaneous localization and mapping. In *Conference on Robot Learning*.

Rosenthal, S.; Biswas, J.; and Veloso, M. M. 2010. An effective personal mobile robot agent through symbiotic human-robot interaction. In *International Conference on Autonomous Agents and Multi-Agent Systems*, 915–922.

Sadat, S. A.; Chutskoff, K.; Jungic, D.; Wawerla, J.; and Vaughan, R. 2014. Feature-rich path planning for robust navigation of MAVs with mono-SLAM. In *IEEE International Conference on Robotics and Automation*, 3870–3875.

SAE On-Road Automated Vehicle Standards Committee. 2014. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *SAE Standards Journal*, 3016: 1–16.

Sternberg, R. J.; and Kolligian Jr., J. 1990. *Competence considered.* Yale University Press.

Triggs, B.; McLauchlan, P. F.; Hartley, R. I.; and Fitzgibbon, A. W. 1999. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, 298–372.

Veloso, M. M.; Biswas, J.; Coltin, B.; and Rosenthal, S. 2015. CoBots: Robust symbiotic autonomous mobile service robots. In *International Joint Conference on Artificial Intelligence*, 4423–4429.