From Shared Control to Symbiosis: A Conflict-Aware Pre-Symbiotic Arbitration Model

Balint Varga
Institute of Control Systems, Karlsruhe Institute of Technology
76131 Karlsruhe, Germany
balint.varga2@kit.edu

Tassilo Ianniello
Institute of Ergonomics, RWTH Aachen
52062 Aachen, Germany
t.ianniello@iaw.rwth-aachen.de

Anja Dannewitz
Fraunhofer FKIE
53343 Wachtberg, Germany
anja.dannewitz@fkie.fraunhofer.de

Frank Flemisch
Institute of Ergonomics, RWTH Aachen
52062 Aachen, Germany
Fraunhofer FKIE, 53343 Wachtberg, Germany

***Abstract*—Human-machine symbiosis is a meta pattern of dynamic interactions between humans and machines, characterized by a subjective oneness between the human and the machine and a high degree of alignment in physically coupled shared control. Reaching and maintaining this symbiotic state is challenging because conflicts can arise from mismatched objectives, internal models, goals, and control actions. Therefore, this paper focuses on pre-symbiotic arbitration: a mediator algorithm is proposed that identifies and resolves conflicts in order to facilitate the transition toward the symbiotic state. We formalize symbiotic state, conflict state, and arbitration, and propose a pattern handler capturing directional opposition and effort mismatch between human and automation inputs. Using this pattern handler, we introduce a conflict-aware mediator model in which automation can arbitrate by goal yielding of its reference to adapt toward the human reference. In a simulation study using 2-DOF shared navigation with goal mismatch, we show that the proposed pattern handler can successfully model the goal mismatch. In addition, the conflict-aware mediator algorithm can resolve this conflict by arbitration of the automation goal, which leads to convergence of the human and the automation goals.**

***Index Terms*—Human-Machine Symbiosis, Shared Control, Pre-symbiotic Arbitration, Conflict-Aware Mediation, Goal Yielding, Human-Robot Interaction, Cooperative Control**

I. Introduction

Human-machine symbiosis - a meta pattern of dynamic interactions characterized by subjective "oneness" and a high alignment in physically coupled shared control - is often treated as the ultimate goal of cooperative human-machine system design [1]–[3]. Yet symbiosis is fragile: even under aligned high-level objectives, mismatches in internal models, intent, or control effort can trigger conflicts that degrade performance, trust, and safety [4]. To support reliable transitions toward symbiosis, we consider a pattern handler and a mediation algorithm that detect opposition and goal alignment and mediate opposition by transforming conflict into a stimulus for symbiosis.

In shared-control settings, one of the central design challenges is how to allocate control and task authority to enable effective cooperation under uncertainty, disturbances, and changing human intent [5]. Cooperative shared control therefore requires a continuously adjusted balance between human and automation in terms of authority, ability, responsibility, and control [6]. A hierarchical view of shared-control conflicts highlights typical sources across intent, information gathering/processing, decision-making, and action implementation [7].

At first glance, conflict may not seem to belong in human-machine symbiosis, often described as the subjective "oneness" between human and machine. A more systemic view suggests the opposite: conflict detection and resolution are foundational prerequisites for the formation and maintenance of symbiosis [4]. Humans and machines begin as autonomous partners; symbiosis emerges through mutual adaptation. Yet differences in capabilities, internal models, or references can cause goal and action conflicts despite shared high-level aims [8], [9]. Because this interaction is dynamic and sensitive to external changes and human state shifts (intent drift, fatigue, mode confusion), systems must detect, analyze, and resolve conflicts in real time.

The central challenge is that the conflict detection is scientifically nontrivial because only a subset of conflicts may be specified a priori. While some are explicitly modeled ("known knowns"), others can only be anticipated as estimable uncertainties before they occur ("known unknowns"), and the most dangerous cases may arise from unmodeled interactions and emergent effects between subsystems and the environment ("unknown unknowns"), becoming understandable only a posteriori. This motivates design methods that (i) define conflict in measurable terms, (ii) tie conflict metrics to arbitration and authority allocation, and (iii) support transitions toward symbiosis through explicit rather than implicit interaction. In particular, a central objective is to turn "known unknowns" into "known knowns" by systematic modeling, monitoring, and iterative refinement.

This paper presents a system-theoretic model of *pre-symbiotic arbitration* – a mechanism that detects and resolves emerging conflicts by adapting agent goals toward symbiosis. We (1) formalize symbiotic and conflict states and the system-theoretic definition of pre-symbiotic arbitration, (2) propose input-level conflict metrics, and (3) an initial validation in a 2-DOF shared navigation study with goal mismatch.

II. Related Work: Arbitration and Conflict in Shared Control Toward Symbiosis

Human-machine arbitration negotiates joint control [10]. As automation gains authority, conflicts arise when human and machine inputs diverge. Therefore, this overview examines arbitration for resolving such conflicts.

A. Ergonomic and Human-Factors Perspective

Consistent mental models enable a dynamic balance between these two agents [6]. However, in reality, there are often conflicts between these mental models. Possible sources of conflict in shared control situations between the human and the machine have in the past been identified as intent, information gathering, information processing, decision-making, and action implementation [7]. Arbitration of conflicts is one key aspect that needs to be included in a cooperative control design. It acts as a structured negotiation tool between the agents to support a unified decision-making process [10]. The resulting state between a human and a machine can be defined as symbiotic - as the highest form of interaction, resulting in the unity of both entities [3]. This unified state can be divided into reversible and non-reversible symbiosis, where non-reversible refers to agents not being able to live and act independently [11]. Within this work, the focus is on reversible symbiosis.

B. Control-Theoretical Models and Conflict Resolutions

Control-theoretic arbitration frameworks increasingly treat human-machine interaction as a multi-layered, dynamic optimization problem [12], which models these interactions as a differential game [13]. Model Predictive Control provides a complementary framework by embedding driver models and safety constraints, enabling minimal-intervention authority management through safe driving envelopes [14] and neuromuscular-model-informed haptic controllers [15]. For scenarios with ambiguous human intent, POMDP-based shared autonomy optimizes over the full distribution of possible goals rather than committing to a single prediction [16], while fuzzy-logic arbitration offers an interpretable alternative for continuous authority allocation [17]. Safety-critical extensions include control barrier functions for arbitration with formal safety certificates [18]. Key open challenges include the integration of computational trust models with game-theoretic controllers, scalable N-player formulations [19], and principled treatment of multi-source uncertainty in arbitration [5], [20]. Within this work, we focus on the input-level conflict metrics and goal-yielding arbitration.

C. Research Gap

There is a research gap in real-time, input-level conflict metrics that directly quantify mismatch between human and automation and that are explicitly linked to the transition toward symbiosis. Current approaches lack a system-theoretic grounding that maps input conflicts to high-level symbiotic states.

III. System-Theoretic Framing of Pre-Symbiotic Arbitration

We define a conflict state as a meta pattern in which the human-machine system exhibits sustained misalignment that prevents symbiotic interaction. Such states typically arise from misalignment in goals, constraints, intent estimates, or control commands between partners, degrading joint performance. Because misalignment can emerge even after coordination has formed [21], conflict episodes can also occur within an otherwise symbiotic interaction and may temporarily reduce the degree of symbiosis. Through arbitration, agents can resolve these conflicts. When holistically viewing a system three states can be identified: *non-symbiotic*, *pre-symbiotic*, and *symbiotic*. In many applications, interaction begins in a non-symbiotic state, i.e., before coordination mechanisms and shared representations have formed. At this point, both entities have an individual understanding of a situation but have not yet communicated their goals with each other. A lack of willingness to cooperate or incompatible goals are signs of a non-symbiotic state. It is the pre-symbiotic state in which a handshake takes place and the two agents gather information on each other. They show initiative to reduce conflicts and use arbitration to improve the goal-reaching process. The word "pre" implies that a symbiotic state is imminent, which can be viewed as an overall common goal being reached if conflict reduction is successful. Eventually, a symbiotic state is reached, which can either remain stable or be lost. The loss of a symbiotic state can have different reasons, e.g., misaligned or changing goals between partners or external influences. The intentions of the agents and their ability to reconnect define whether the new system state is pre- or non-symbiotic. Without bilateral engagement of the agents, a symbiotic state cannot be reached. The following working definition is proposed to define the term **pre-symbiotic arbitration**:
"Pre-symbiotic arbitration is the mediating mechanism that identifies emerging goal conflicts and resolves them by arbitrating and allocating control and task authority to enhance the interaction between human and automation and achieve symbiosis".

In the following, a framework is proposed to explain the pre-symbiotic arbitration process between two agents (see Fig. 1). In this industrial example, a human and an automation are jointly controlling a machine - in this case, a robot in a production plant. The two agents interact with each other while pursuing their goals. Their overall goal of safely and efficiently controlling the machine overlap. However, on a smaller scale, the realization of the goals may differ. Without explicitly communicating their actions, their goals and states can be

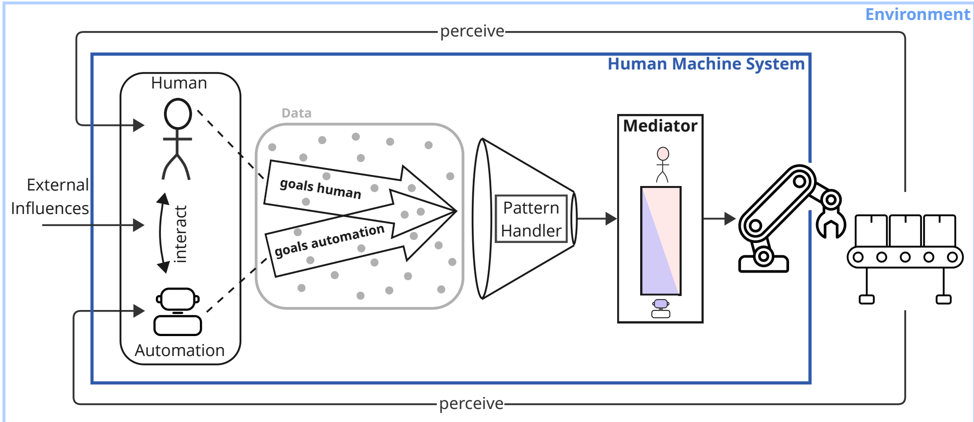

Fig. 1: Example of a Pre-Symbiotic Arbitration Framework. Icons by Doodle Icons, Lilik Sofiyanti and Marcus DeClarke (Noun Project)

derived by processing the data, presented as dots in Fig. 1. Data can include information about the human, such as vital signs, time into shift or individual trajectory preferences of the robot. Data about the automation can be mathematical models of trajectories which may differ from the human trajectories. In addition to those two data sources, environmental data may influence the arbitration process. Data in this category might be sensor data about the proximity of living beings or objects to the robot.

As the human and the automation give their inputs, the data is evaluated in the form of patterns. This happens in the instance of the pattern handler. The pattern handler can be understood as a hub that aggregates information in the form of patterns to share them with the mediator [22]. Thereby, the predetermined patterns are given a proximity of being correct. The mediator then uses the provided information to facilitate an adequate control distribution between human and automation. Consequently, a symbiosis can be understood as a meta pattern [11]. An example would be a human who has worked with the system for a longer period of time and has repeatedly made a mistake, causing a collision with an object in the path of his chosen trajectory. As the mistake is frequently re-occurring in the same location, the pattern handler identifies an increased risk of danger. The preferred trajectory of the automation in this location is collision-free. This trajectory data is documented and communicated to the mediator. The mediator then orchestrates the decision making process. It does not decide on an action itself, as this is previously done by the agents and their input; it much rather aligns weight to the decisions made, based on the provided patterns, to leverage the final action as the most efficient result for the whole system. As a result, the trajectory, goals and control authority become updated, avoiding a collision, but encouraging symbiosis.

The framework shows an iterative and dynamic process, which is an essential element of pre-symbiotic arbitration. Both agents, the human and the automation, perceive information about the environment and the consequences of their actions, which is then used as input to further improve their understanding of these actions. In addition, external influences, e.g., software updates of the automation or colleagues talking to the human, might influence the process of reaching symbiosis as well.

The potential of the process of reaching symbiosis can be beneficial for training similar systems or improving the stability of symbiotic states. One possible facilitator may be a digital shadow, representing individual human behavior, habits, or goals. The pattern handler can adapt patterns according to individual shadows, to enable a more personalized interaction between the human and the machine, due to a more aligned understanding of the corresponding goals. In addition, digital shadows could be transferred between human-machine systems, to allow the automation algorithms to adapt faster to a human agent. This machine-learning-based adaptability will further refine the interaction parameters to resolve conflict states in a more efficient manner. Additionally, this scheme would use data to gather information, which can then be hardened to knowledge about a situation or, in general, human agents as users, and finally be reused in different scenarios, resulting in wisdom.

IV. Mathematical Pre-symbiotic Arbitration Model With Conflict-Aware Goal Yielding

Building on the framework from Section III, this section introduces a mathematical model of pre-symbiotic arbitration: a conflict-aware mediator that detects input disagreements and resolves them by adaptively yielding the automation's goal toward the human's reference.

A. Human-Machine System Model

We consider a linear discrete-time plant modelling the human-machine shared control system within their environment:

$$\boldsymbol{x}(k+1) = \mathbf{A}\boldsymbol{x}(k) + \mathbf{B}^{\mathrm{a}}\boldsymbol{u}^{\mathrm{a}}(k) + \mathbf{B}^{\mathrm{h}}\boldsymbol{u}^{\mathrm{h}}(k), \tag{1}$$

where system state is $\boldsymbol{x}(k) \in \mathbb{R}^n$, and the human and the automation have the control inputs $\boldsymbol{u}^{\mathrm{h}}(k) \in \mathbb{R}^m$ and $\boldsymbol{u}^{\mathrm{a}}(k) \in \mathbb{R}^m$, respectively. The control inputs are obtained by the optimization problems of the two agents:

$$J^i = \sum_{k=0}^{\infty} \Big((\boldsymbol{x}(k) - \boldsymbol{r}^i)^\top \mathbf{Q}^i (\boldsymbol{x}(k) - \boldsymbol{r}^i) + \boldsymbol{u}_k^\top \mathbf{R}^i \boldsymbol{u}_k \Big), \tag{2}$$

where $i \in \{h, a\}$, with $\mathbf{Q}^i \succeq 0$, $\mathbf{R}^i \succ 0$, and $\boldsymbol{r}^{\mathrm{h}}(k) \in \mathbb{R}^n$ denote the human reference (goal state) at time step k, and let $\boldsymbol{r}^{\mathrm{a}}(k) \in \mathbb{R}^n$ denote the automation's current goal. Note that $\boldsymbol{r}^{\mathrm{h}} \neq \boldsymbol{r}^{\mathrm{a}}$. This allows modeling goal mismatches that can lead to conflicts. This "conflict-by-design" setting differs from prior work on uncertainty-induced conflict [5], [20]. In case of linear dynamics (1), both agents yield linear optimal feedback laws:

$$\boldsymbol{u}^i(k) = -\mathbf{K}^i\big(\boldsymbol{x}(k) - \boldsymbol{r}^i\big), \quad i \in \{\mathrm{h}, \mathrm{a}\}, \tag{3}$$

where $\mathbf{K}^i$ are the feedback gains obtained by solving the discrete-time Riccati equation. This setup is a classical shared control scenario with two optimal controllers having slightly different goals, which can lead to conflicts in their control inputs.

B. Arbitration Model of the Pattern Handler

To model conflict in shared control, we introduce an arbitration mechanism that merges inputs from the human and the automation. This model subsequently can serve to mediate conflicts and to achieve symbiosis formation by explicitly identifying mismatched decisions and actions, and by adapting the allocation of authority in real time. The control inputs of the two agents are blended as a convex combination:

$$\boldsymbol{u}(k) \;=\; \alpha_k\, \boldsymbol{u}^{\mathrm{h}}(k) + \big(1 - \alpha_k\big)\, \boldsymbol{u}^{\mathrm{a}}(k), \;\; \alpha_k \in [0, 1]. \tag{4}$$

In this paper, we consider a case in which α_k is constant. In general settings, α_k can be state/conflict dependent, enabling more advanced adaptation strategies, see e.g. [23].

To model input-level conflict, we define metrics that quantify disagreement between the agents' candidate inputs at time step k. The pattern handler measures both directional opposition (aligned vs. opposite control actions) and effort mismatch (similar vs. divergent magnitudes) to detect and quantify conflicts. We define two bounded metrics at time step k:

$$\phi^{\mathrm{opp}}_{\mathrm{raw}}(k) = \frac{1}{2}\left(1 - \frac{\boldsymbol{u}^{\mathrm{h}}(k)^\top \boldsymbol{u}^{\mathrm{a}}(k)}{\|\boldsymbol{u}^{\mathrm{h}}(k)\|\, \|\boldsymbol{u}^{\mathrm{a}}(k)\| + \varepsilon}\right) \in [0, 1], \tag{5}$$

$$\phi^{\mathrm{mag}}(k) = \frac{\big|\|\boldsymbol{u}^{\mathrm{h}}(k)\| - \|\boldsymbol{u}^{\mathrm{a}}(k)\|\big|}{\|\boldsymbol{u}^{\mathrm{h}}(k)\| + \|\boldsymbol{u}^{\mathrm{a}}(k)\| + \varepsilon} \in [0, 1]. \tag{6}$$

Directional opposition (aligned$\to$ 0, opposite$\to$ 1) and effort mismatch (similar$\to$ 0, dominant$\to$ 1) are measured; $\varepsilon > 0$ avoids division by zero. Note that (5) and (6) are extensions of [20] by including a directional opposition metric in addition to the effort mismatch, allowing them to capture a wider range of conflict patterns.

If both inputs are near zero, cosine-based quantities can become numerically unstable. Therefore, we include the total input energy $E_{\mathrm{input}}(k) = \|\boldsymbol{u}^{\mathrm{h}}(k)\|^2 + \|\boldsymbol{u}^{\mathrm{a}}(k)\|^2$ in the conflict logit as a normalization term, which reduces the impact of noisy disagreement when both inputs are small. The following smoothing gate

$$g^{\mathrm{dir}}(k) = \frac{E_{\mathrm{input}}(k)}{E_{\mathrm{input}}(k) + \delta}, \quad \delta > 0, \tag{7}$$

is applied to the directional opposition metric, yielding the gated opposition feature $\phi^{\mathrm{opp}}(k) = g^{\mathrm{dir}}(k)\, \phi^{\mathrm{opp}}_{\mathrm{raw}}(k)$. Define the feature vector $\boldsymbol{\phi}(k) = [\phi^{\mathrm{opp}}(k),\ \phi^{\mathrm{mag}}(k)]^\top$ and weights $\boldsymbol{w} = [w_{\mathrm{opp}},\ w_{\mathrm{mag}}]^\top$ with $w_{\mathrm{opp}}, w_{\mathrm{mag}} \geq 0$. The *conflict logit* is

$$\begin{aligned} s(k) =& \frac{1}{E_{\mathrm{input}}(k) + \varepsilon} + \boldsymbol{w}^\top \boldsymbol{\phi}(k) \\ =& \frac{1}{E_{\mathrm{input}}(k) + \varepsilon} + w_{\mathrm{opp}} \phi^{\mathrm{opp}}(k) + w_{\mathrm{mag}} \phi^{\mathrm{mag}}(k), \end{aligned} \tag{8}$$

and the normalized conflict index is

$$c(k) = \frac{1}{1 + \exp(-s(k))} \in (0, 1). \tag{9}$$

C. Goal-Yielding Model of the Mediator

Another possibility to resolve conflicts is to adapt the goals of the agents. To model this "yielding", we introduce a concession state $\lambda(k) \in [0, 1]$ that modulates the automation's goal reference $\boldsymbol{r}^{\mathrm{a}}(k)$ toward the human reference $\boldsymbol{r}^{\mathrm{h}}(k)$. In this work, the human reference is assumed to be directly available to the automation (e.g., communicated explicitly, provided by an interface, or known by design in simulation). Hence, no intent inference or reference estimation is required[1].

To model goal adaptation as a negotiation mechanism of the automation, its goal is updated by yielding toward the human reference using a convex combination:

$$\boldsymbol{r}^{\mathrm{a}}(k+1) = (1 - \lambda(k))\, \boldsymbol{r}^{\mathrm{a}}(k) + \lambda(k) \boldsymbol{r}^{\mathrm{h}}(k). \tag{10}$$

The scalar $\lambda(k)$ is interpreted as the goal yielding function: $\lambda(k) = 0$ keeps the automation goal unchanged, while $\lambda(k) = 1$ makes the automation fully adopt the human reference in one step. Intermediate values implement gradual convergence and allow tuning the time scale of goal adaptation.

To connect yielding to the interaction quality, we choose $\lambda(k)$ as a monotone function of a conflict index $c(k) \in (0, 1)$, where higher values indicate stronger disagreement between the human and automation at the inputs. The bounded mapping is

$$\lambda(k) \;=\; \min\{1, \max\{0, \eta\, c(k)\}\}, \;\; \eta > 0, \tag{11}$$

Thus, when the interaction is more symbiotic (low c_k), the automation yields slowly, whereas sustained conflict increases $\lambda(k)$ and accelerates convergence of $\boldsymbol{r}^{\mathrm{a}}(k)$ toward the human reference. Finally, the automation input $\boldsymbol{u}^{\mathrm{a}}(k)$ is computed with respect to its arbitrated goal $\boldsymbol{r}^{\mathrm{a}}(k)$, while the human reference $\boldsymbol{r}^{\mathrm{h}}$ remains constant.

Equations (10) and (11) provide a baseline for studying how conflict-driven goal yielding can move a shared-control system toward higher alignment when the human reference is known to the automation. This setup isolates the effect of the yielding law itself by keeping the arbitration and controller structure fixed, so changes in behavior can be attributed to the conflict feedback. The summary of the overall process is given in Algorithm 1.

[1] In some cases, the human's reference is not directly available and must be estimated, see e.g., [24]

Algorithm 1 Pre-Symbiotic Conflict-Aware Arbitration

Initialize $\boldsymbol{x}(0)$, $\boldsymbol{r}^{\mathrm{a}}(0)$, $\boldsymbol{r}^{\mathrm{h}}(0)$, and parameters.
for $k = 0, 1, 2, \ldots$ **do**
  Agent
    Compute candidate inputs $\boldsymbol{u}^{\mathrm{h}}(k)$ and $\boldsymbol{u}^{\mathrm{a}}(k)$ via (3)
  Pattern handler
    Compute $\phi^{\mathrm{opp}}_{\mathrm{raw}}(k)$ and $\phi^{\mathrm{mag}}(k)$ via (5)–(6)
    Compute $g^{\mathrm{dir}}(k)$ via (7)
    Set $\phi^{\mathrm{opp}}(k) \leftarrow g^{\mathrm{dir}}(k)\, \phi^{\mathrm{opp}}_{\mathrm{raw}}(k)$
    Compute $s(k)$ via (8) and $c(k)$ via (9)
  Mediator
    Set $\lambda(k) \leftarrow \min\{1, \max\{0, \eta c(k)\}\}$ via (11)
    Update $\boldsymbol{r}^{\mathrm{a}}(k+1)$ via (10)
    Apply $\boldsymbol{u}(k) \leftarrow \alpha\, \boldsymbol{u}^{\mathrm{h}}(k) + (1-\alpha)\, \boldsymbol{u}^{\mathrm{a}}(k)$ via (4)
  Environment
    Propagate $\boldsymbol{x}(k+1)$ via (1)
end for

V. Simulation Study: Conflict-to-Symbiosis Transition in 2-DOF Shared Navigation

To illustrate conflict-aware arbitration and symbiosis formation, we design a simulation study of 2-DOF shared navigation with conflicting goals. The plant is a 2-DOF point mass with discrete-time double-integrator dynamics, where $\boldsymbol{x}(k) = [p_x(k), p_y(k), v_x(k), v_y(k)]^\top$ contains position and velocity, $\boldsymbol{u}(k) = [a_x(k), a_y(k)]^\top$ is the acceleration input, and $\Delta t = 0.025$ s is the sampling time.

The parameters of the cost functions (2) are set to $\mathbf{Q}^{\mathrm{h}} = \mathrm{diag}([20.0, 20.0, 12.0, 12.0])$, for the human and $\mathbf{Q}^{\mathrm{a}} = \mathrm{diag}([30.0, 30.0, 3.0, 3.0])$ for the automation. The input costs are $\mathbf{R}^{\mathrm{a}} = \mathbf{R}^{\mathrm{h}} = \boldsymbol{I}_2$. Solving the coupled Riccati equations leads to feedback gains

$$\mathbf{K}^{\mathrm{h}} = \begin{bmatrix} 7.3 & 0 & 4.2 & 0 \\ 0 & 7.3 & 0 & 4.2 \end{bmatrix}, \mathbf{K}^{\mathrm{a}} = \begin{bmatrix} 5.2 & 0 & 3.6 & 0 \\ 0 & 5.2 & 0 & 3.6 \end{bmatrix}.$$

The references are set to $\boldsymbol{r}^{\mathrm{h}} = [5.0, 0.0, 0.0, 0.0]^\top$ for the human and $\boldsymbol{r}^{\mathrm{a},0} = [5.0, 0.6, 0.0, 0.0]^\top$ for the automation, creating a vertical goal mismatch. The parameters of the conflict logit are set to $w_{\mathrm{opp}} = 4.0$, $w_{\mathrm{mag}} = 4.0$, $\varepsilon = 0.01$, and $\delta = 0.01$. The yielding rate is set to $\eta = 0.02$. The arbitration blending weight is constant $\alpha = 0.5$.

A. Results and Discussion

In the simulations, two scenarios are compared: one with conflict-aware goal yielding (Algorithm 1) and one without yielding (i.e., $\lambda(k) \equiv 0$).

In the non-yielding case, the trajectory results in a persistent conflict state between the two goals, maintaining a remaining control input; see Figs. 2 and 3. As neither agent adapts its goal, the conflict state remains high, as shown in Fig. 4. The conflict index converges to a high value, indicating sustained disagreement between the agents' inputs due to the goal mismatch.

In the yielding case, the automation's goal converges toward the human's goal, see Fig. 2, leading to aligned and vanishing control inputs, see Fig. 5. The conflict state is reduced to zero, as indicated in Fig. 6. The conflict index converges to zero, reflecting the resolution of disagreement as the automation yields its goal toward the human's reference.

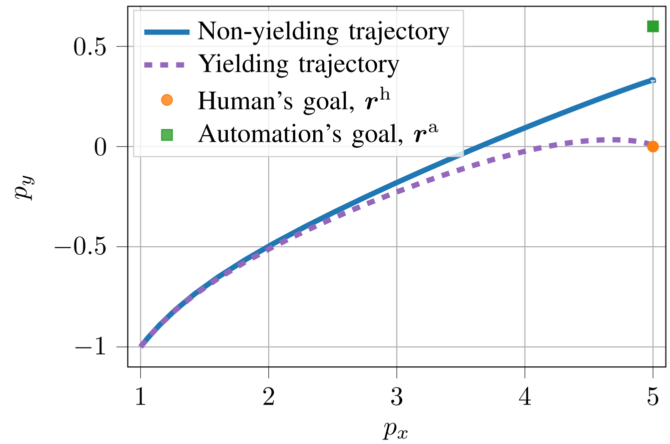

Fig. 2: The resulting trajectory in 2-DOF shared navigation with and without yielding.

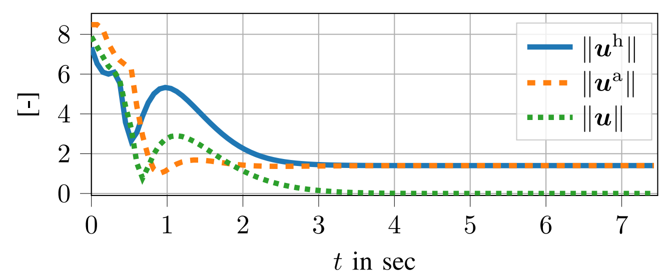

Fig. 3: The norms of the controls in the shared navigation without yielding.

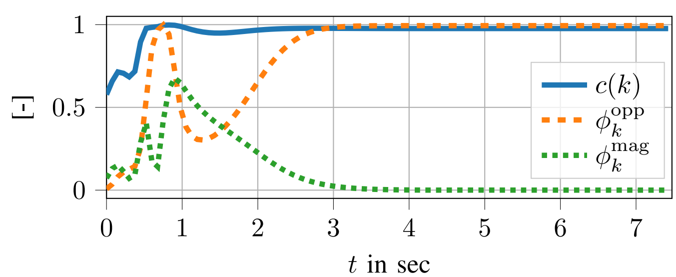

Fig. 4: The measures of the conflict states in the shared navigation without yielding.

The conflict-aware yielding mechanism reduces goal mismatch and drives the system toward a low-conflict (symbiotic) state. In contrast, a fixed goal mismatch yields persistent conflict. This example highlights how conflict measures can guide arbitration toward symbiosis and can be extended to settings with uncertainty and partial information.

VI. Conclusion and Future Work

This paper presented a system-theoretic framing of pre-symbiotic arbitration as a mediator. We synthesized representative arbitration mechanisms in relation to symbiosis dimen-

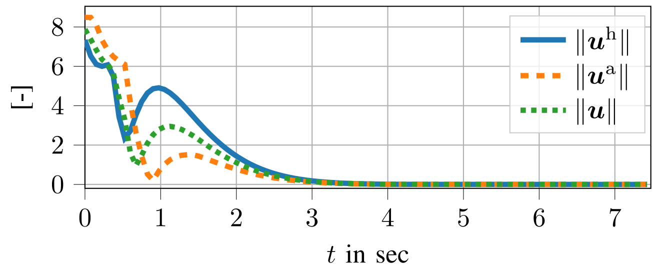

Fig. 5: The norms of the controls in the shared navigation with yielding.

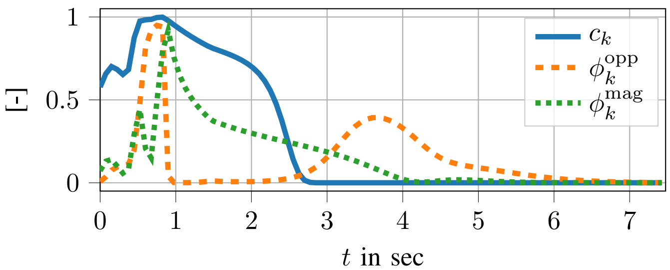

Fig. 6: The measures of the conflict states in the shared navigation with yielding.

sions and conflict emergence, illustrated conflict emergence from slightly mismatched optimal goals, and evaluated arbitration using conflict-state measures aligned with cooperation quality. Future work will extend the mathematical model to include uncertainty-aware arbitration and more complex goal negotiation strategies, and will validate the framework in human-subject experiments with a physical shared-control setup. The overall vision is to understand and develop arbitration strategies that can guide human-machine systems toward symbiosis.

References

[1] J. C. R. Licklider, "Man-computer symbiosis," *IRE Transactions on Human Factors in Electronics*, vol. HFE-1, no. 1, pp. 4–11, 1960.

[2] J. Y. Chen and M. J. Barnes, "Human–robot interaction," *Handbook of human factors and ergonomics*, pp. 1121–1142, 2021.

[3] J. Inga, M. Ruess, J. H. Robens, T. Nelius, S. Rothfuß, S. Kille, P. Dahlinger, A. Lindenmann, R. Thomaschke, G. Neumann, S. Matthiesen, S. Hohmann, and A. Kiesel, "Human-machine symbiosis: A multivariate perspective for physically coupled human-machine systems," *International Journal of Human–Computer Studies*, vol. 170, p. 102926, 2023.

[4] F. O. Flemisch, M.-P. Pacaux-Lemoine, F. Vanderhaegen, M. Itoh, Y. Saito, N. Herzberger, J. Wasser, E. Grislin, and M. Baltzer, "Conflicts in human-machine systems as an intersection of bio- and technosphere: Cooperation and interaction patterns for human and machine interference and conflict resolution," in *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, 2020, pp. 1–6.

[5] S. Jiang, Y.-B. Zhao, Y. Kang, F. Xie, and Y.-S. Zhao, "Una-sac: An uncertainty-aware nonlinear arbitration method for human-ai shared control," *IEEE Transactions on Artificial Intelligence*, 2026, early Access.

[6] F. Flemisch, M. Heesen, T. Hesse, J. Kelsch, A. Schieben, and J. Beller, "Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations," *Cognition, Technology & Work*, vol. 14, no. 1, pp. 3–18, 2012.

[7] M. Itoh, F. Flemisch, and D. Abbink, "A hierarchical framework to analyze shared control conflicts between human and machine," *IFAC-PapersOnLine*, vol. 49, no. 19, pp. 96–101, 2016.

[8] M. Grobbel, B. Varga, and S. Hohmann, "Shared telemanipulation with vr controllers in an anti slosh scenario," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2023, pp. 4405–4410.

[9] J. Schneider, B. Varga, and S. Hohmann, "Cooperative trajectory planning: Principles for human-machine system design on trajectory level," in *at-Automatisierungstechnik*, vol. 72, no. 12, pp. 1121–1129.

[10] J. Kelsch, F. Flemisch, C. Löper, A. Schieben, and J. Schindler, "Links oder rechts, schneller oder langsamer? Grundlegende Fragestellungen beim Cognitive Systems Engineering von hochautomatisierter Fahrzeugführung," In Hrsg. Grandt, M; Bauch, A.: *Cognitive Systems Engineering i.d. Fahrzeug- und Prozessführung*; 48. FAS Anthropotechnik, Karlsruhe, 2006.

[11] F. Flemisch and M. Baltzer, "Are rider-horse or centaurs intelligent human systems integration? First sketch of reversible and non-reversible human technology/machine/ai symbiosis," *iHSI Int. Conference on Intelligent Human Systems Integration*, vol. 22, no. 22, Venice, Italy, 2022.

[12] D. P. Losey, C. G. McDonald, E. Battaglia, and M. K. O'Malley, "A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction," *ASME Applied Mechanics Reviews*, vol. 70, no. 1, p. 010804, 2018.

[13] P. Franceschi, N. Pedrocchi, and M. Beschi, "Human–robot role arbitration via differential game theory," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 4, pp. 5953–5968, 2024.

[14] S. M. Erlien, S. Fujita, and J. C. Gerdes, "Shared steering control using safe envelopes for obstacle avoidance and vehicle stability," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 441–451, 2016.

[15] A. M. Lazcano, T. Niu, H. C. G. Akutain, D. J. Cole, and B. Shyrokau, "MPC-based haptic shared steering system: A driver modeling approach for symbiotic driving," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 3, pp. 1201–1211, 2021.

[16] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. A. Bagnell, "Shared autonomy via hindsight optimization for teleoperation and teaming," *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 717–742, 2018.

[17] X. Xing, W. Li, S. Yuan, and Y. Li, "Fuzzy logic-based arbitration for shared control in continuous human–robot collaboration," *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 7, pp. 3979–3991, 2024.

[18] M. Y. Uzun and Y. Yildiz, "Arbitration with control barrier functions for safe shared control," *IEEE Control Systems Letters*, vol. 9, pp. 2789–2794, 2025.

[19] B. Varga, J. Inga, M. Lemmer, and S. Hohmann, "Ordinal Potential Differential Games to Model Human-Machine Interaction in Vehicle-Manipulators," in *2021 IEEE Conference on Control Technology and Applications (CCTA)*. San Diego, CA, USA: IEEE, Aug. 2021, pp. 728–734.

[20] S. Li, M. Bowman, and X. Zhang, "A general arbitration model for robust human-robot shared control with multi-source uncertainty modeling," arXiv:2003.05097, 2020.

[21] M. Baltzer, E. Altendorf, S. Meier, and F. Flemisch, "Mediating the interaction between human and automation during the arbitration processes in cooperative guidance and control of highly automated vehicles: Basic concept and first study," In: (ed.) Stanton, N., Landry, S., Bucchianico, G. D., and Vallicelli, A. *Advances in Human Aspects of Transportation Part I*, pp. 439-450. AHFE Conference, 2014.

[22] M. Usai, N. Mandischer, and F. Flemisch, "Pattern handler: Integrating real world and virtual models of human systems patterns to regulate the control distribution in a cooperative automated driving task," in *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2024, pp. 916–923.

[23] C. A. Braun, L. Haide, L. Fischer, S. Kille, B. Varga, S. Rothfuss, and S. Hohmann, "Using a collaborative robotic arm as human-machine interface: System setup and application to pose control tasks," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 12 464–12 470.

[24] J. Schneider, B. Varga, and S. Hohmann, "Consideration of communication in human–machine interaction for cooperative trajectory planning," in *Frontiers in Robotics and AI* vol. 12, p. 1568402.