1 Additional Information about the Experimental Evaluation

1.1 Simulation Setup and Parameters

We considered a normative multi-agent system comprising actor agents and observer agents. We assumed a fixed number of actors (100) and evaluate the learning experience of a single observer observing the behaviors of these actors. During each simulation run, agents (actors and observers) sense and act (i.e., perform behaviors) repeatedly for number of cycles, represented by the *lifetime* parameter. We assumed that at the start of the simulation, the norm system had reached some stable normative state and certain behaviors come assigned with a normative status - Forbidden, Obligatory, or Optional, which we assigned randomly. This served as a clear **ground-truth** with which to evaluate the performance of the proposed approach. The actors were aware of the normative status of the behaviors and their choice of performing a particular behavior during a cycle was governed by the normative status of that behavior and the actor's own *compliance rate*, a controlled experimental parameter for Type A uncertainty. We assumed that the observer had full knowledge about the possible set of behaviors that an actor could display, but none about their normative status.

At every time step, the observer observed the set of behaviors during an interaction encounter between two randomly selected actors. The observers could also additionally observe if a particular behavior was sanctioned. The sanctioning, itself, was performed not by the actors or observer, but externally by the system, per the experimental parameter *sanction rate*. That is, in each cycle, every actor was in a particular context, and performed or omitted behaviors within that context. If the actor performed a forbidden behavior or omitted an obligatory one, there was a chance that the environment would sanction this actor. We did not model the effect on the actor of this sanction. Instead we focused on what the observer learned from it. The observer performed a hard-correct based on the observed sanction and then updated its belief. Sensor reliability was controlled with a *sensor reliability* parameter, which featured as a DS-theoretic mass assignment of less than 1 for an observation. Observational ambiguity was controlled with a *ambiguity rate* parameter (Type C uncertainty), that governed the proportion of missing data, i.e., the number of ϵ entries in the data BoE.

We made certain simplifying assumptions in this setup to focus on our core claims. A more sophisticated simulation might be designed to address many additional research questions and might entail agents that are both actors and observers moving through multiple contexts. Sanctioning might be another behavior performed by the agents themselves. Of course, then the agents would need to also be able to distinguish sanctioning from a non-sanctioning behavior. Sanctioned actors could even become more compliant. These are interesting questions that is the subject of future work and well beyond the scope of this current work.

To summarize, each of the different forms of uncertainty was assigned an independent parameter that we varied. The specific values of the parameters used in our experiments are outlined as relevant below.

1.2 Parameter Settings for the Convergence Behavior Experiments

We first study how beliefs and plausibilities change over the course of a simulation run. We assume complete ignorance ($[0, 1]$ uncertainty interval) for each behavior at the start. In this experiment, we compute converge values for these uncertainty intervals through the simulation under different experimental conditions – compliance, sanction rate and ambiguity. For each of the simulation runs, we considered three behaviors and assigned them a different normative status of forbidden, optional or obligatory.

For a particular behavior, the *compliance rate* for each actor was obtained by sampling from a Gaussian distribution centered at either 0.0 (for forbidden behaviors), 1.0 (for obligatory behaviors) and 0.5 (for optional behaviors). Compliance probability was varied by increasing (low compliance) or decreasing (high compliance) the standard deviation. We varied the sanction rate between 1.0 (high sanction rate) and 0.1 (low sanction rate) to provide a probability that a particular behavior was going to be sanctioned. High ambiguity rates suggests that a large amount of missing information about whether a behavior was performed or not.

1.3 Parameter Settings for the Performance Evaluation

We next study the performance of the norm learning as measured by precision and recall of the normative statuses of the behaviors. We performed a larger scale simulation in which we varied lifetimes – 25, 50, 100 and 500, compliance rates (low/high), sanction rates (low/high), ambiguity rates (low/high), sensor reliability rates (low/high) and number of norms – 3, 5, 7, 10, 12. We considered a 1000 random seeds allowing us a total of 320,000 separate runs of each simulation scenario. Note each simulation scenario comprises a lifetime of observations of an observer of 100 actors. We calculated the precision and recall for the multi-class classification problem of deciding if a behavior was forbidden, obligatory or optional.