First off, I would like to thank the reviewers for their helpful feedback. The reviewers agree that this work provides a novel way of looking at the expressivity of GNNs and establishes insightful results for the important problem of graph isomorphism. Following their suggestions, the paper's exposition will be improved by: 1) providing further (graphical) intuition on the concept of a protocol; 2) unifying Theorems 3.1-3.3 and Prop. 3.1 into a main theorem; and 3) by expliciting the differences with [33] and the 1-WL test (discussed below). I will also address any other minor comments that are not discussed here due to space limitations.

R1: Why is the learnability of the isomorphism class (f_{iso}) important? 1) f_{iso} is a good proxy for graph classification: due to MLP universality, a GNN that solves f_{iso} is sufficiently powerful to solve any graph classification problem on the same graph distribution (i.e., irrespective of how the classes are assigned). 2) The bounds also apply to any GI testing method (like [31]) that compares graphs by means of some invariant representation (see Sec.3:129-133). GI testing is a subject of intense study within the GNN community [11, 22, 31, 25].

R1 & R2: Anonymity, and improvement over 1-WL bounds for trees. The proposed bounds apply *both* to anonymous and non-anonymous MPNN. The tree distribution was chosen purposefully to demonstrate that the bounds are also relevant for the anonymous case. As R1/R2 mentioned, it is known that 1-WL can recognize v-node trees in v iterations, simply because there exists a tree of diameter v. However, since MPNN is equivalent to 1-WL only when the former is built using injective aggregation functions (i.e., of unbounded width), the equivalence does not imply a relevant lower bound on the width/message-size/global-state-size of MPNN. Further, the communication complexity theoretic bound is tighter and more refined: e.g., it asserts that one needs $\Omega(v)$ capacity *in expectation*, even though the average tree has $O(\sqrt{v})$ diameter (and thus 1-WL would require depth=$\Omega(\sqrt{v})$ in expectation).

R1: Experiments: why were they set-up in this manner and are they small-scale? The experiments were intended as a verification of the bounds and thus were constructed to closely match the studied setting. Yet, since the considered distributions form a (size-able) subset of all possible graphs/trees on n nodes, the demonstrated impossibility results will similarly hold for the full distribution. The hyper-parameters $v = n/2, \tau = 1, w \leq 16$ were selected to illustrate the bounds for small n—the bounds also hold for different settings of v, τ, w, but then a larger n would be needed to demonstrate the dependency (rendering the experiments more lengthy). Overall, the experiments considered ∼36.5k different graphs and 420 different MPNN, lasting more than 672 GPU hours. In terms of these metrics, I would not consider them small-scale.

R2: Communication complexity (CC) and coding theory. The applicability of Shannon's theorem (proof of Lemma B.2) arises when shifting from a worst-case complexity definition (studied in CC) to an expected one (defined here). I believe that this is the reason why these connections were not exploited previously.

R3: Relation to [33]. There are four main differences (beyond that [33] did not consider graph isomorphism): 1) *The current paper derives necessity bounds for the expected as well as worst case.* The results of [33] assert that there exists a distribution such that for *some* graph in the distribution dw needs to depend on n — however, the MPNN could still attain 99.99% accuracy without needing to satisfy the dw condition. Differently, the current paper bounds the probability of error over the distribution (see third bullet in Prop. 3.1). Note also that Lemma 2.1 can be used to define a capacity bound for any arbitrary family of graphs (though the bound is tighter when the graphs in question can be jointly partitioned with a cut of at most τ). Further, since anonymous MPNN is oblivious to node ordering, in the anonymous setting the bound is valid as long as each graph can be cut in two pieces of roughly equal size. 2) *In contrast to [33], the bounds here consider the message-size and apply to MPNN with global-state.* 3) *This paper considers graph classification (with readout); [33] considered node classification problems (no readout function).* 4) *Finally, the current paper derives lower bounds by developing a new, technically more involved, and insightful connection to CC.*

R3: Regarding rigor and completeness. 1) *All results were fully proven.* Employing a previous result (Corollary B1) in a proof is standard practice in theoretical papers and does not affect rigor. To help the reader, relevant CC arguments and results are summarized in App. B.1 and B.2. These might not suffice to provide a complete intuition, but they suffice for completeness. To aid the reader further, the camera-ready will include a more in-depth explanation of protocols and communication complexity. 2) The intuition behind the technical constructions is provided by the introduction of Section 3. Further intuition in the main text would require an understanding of App. B, which I believe goes beyond the interests of the casual reader.

R3: Why is a rectangle-based analysis necessary? The argument of the reviewer could be used in a setting where the information needs to be transmitted one-way. However, in MPNN any two subgraphs (parties) arrive to the output by exchanging information over multiple steps/layers (i.e., the transmission is both ways). This renders the approach suggested by the reviewer inapplicable and motivates the need for rectangles/communication complexity: a rectangle represents the uncertainty inherent to each party after every step of the communication exchange (protocol).

R4: Communication capacity. The definition can be found in Definition 2.1 (lines 112-114).