// what observations imply

O1Y→C1Y∨C1B , O1W→C1W∨C1B

O2Y→C2Y∨C2B, O2W→C2W∨C2B

O3Y→C3Y∨C3B, O3W→C3W∨C3B

// labels are wrong

L1Y→¬C1Y, L1W→¬C1W, L1B→¬C1B

L2Y→¬C2Y, L2W→¬C2W, L2B→¬C2B

L3Y→¬C3Y, L3W→¬C3W, L3B→¬C3B

// there is at least 1 box of each color

C1Y∨C1W∨C1B, C2Y∨C2W∨C2B, C3Y∨C3W∨C3B

// no 2 boxes have the same contents

C1Y→¬C2Y∧¬C3Y, C1W→¬C2W∧¬C3W, C1B→¬C2B∧¬C3B

C2Y→¬C1Y∧¬C3Y, C2W→¬C1W∧¬C3W, C2B→¬C1B∧¬C3B

C3Y→¬C2Y∧¬C1Y, C3W→¬C2W∧¬C1W, C3B→¬C2B∧¬C1B

The above propositional rules can be converted to:

# These are observations and labels implies

-O1Y C1Y C1B

-O2W C2W C2B

-O3Y C3Y C3B

-L1W -C1W

-L2Y -C2Y

-L3B -C3B

# These are "at least one box" sentences

C1Y C1W C1B

C2Y C2W C2B

C3Y C3W C3B

# These are "exactly one box" sentences to CNF

-C1Y -C2Y

-C1Y -C3Y

-C1W -C2W

-C1W -C3W

-C1B -C2B

-C1B -C3B

# These are the facts we have known

O1Y

O2W

O3Y

L1W

L2Y

L3B

# We want to prove KB|=C2W

-C2W