

Appendix

Vishvak Murahari¹ Prithvijit Chattopadhyay¹
Dhruv Batra^{1,2} Devi Parikh^{1,2} Abhishek Das¹

¹Georgia Tech ²Facebook AI Research

1 Qualitative Output

The dialog generated by self-talk between different SL model variants is provided in Table. 3 and different RL model variants is provided in Table. 4. We observe the variants with Diverse-Q-BOT tend to generate more diverse, image relevant and fluent dialog.

2 Experiments with Diverse-A-BOT

We also ran experiments where we used the repetition penalty (Eqn. (1) in the main paper) during SL pre-training of A-BOT. In Tables 1 and 2 we report results on the A-BOT retrieval metrics (Das et al., 2017a) and the diversity metrics respectively. In Fig. 1, we provide the performance of various Q-BOT variants paired with this Diverse-A-BOT on the image-guessing task.

We note that retrieval metrics for Diverse-A-BOT are better than vanilla SL: A-BOT. Therefore, this repetition penalty does help significantly during supervised pre-training. However, finetuning the Diverse-A-BOT via RL does not lead to significant improvements.

We note that for the diversity metrics on Q-BOT-A-BOT dialog, self-talk with Diverse-A-BOT does not change the diversity metrics significantly. We observe the same trend in image-guessing performance as well.

3 Model Architecture details

We use a Hierarchical Recurrent Encoder (HRE) for representing dialog context. In this encoder the image representation is concatenated with every question word when fed to the LSTM. We then encode each QA-pair in the dialog history with another LSTM with shared weights. The image-question representation, computed for

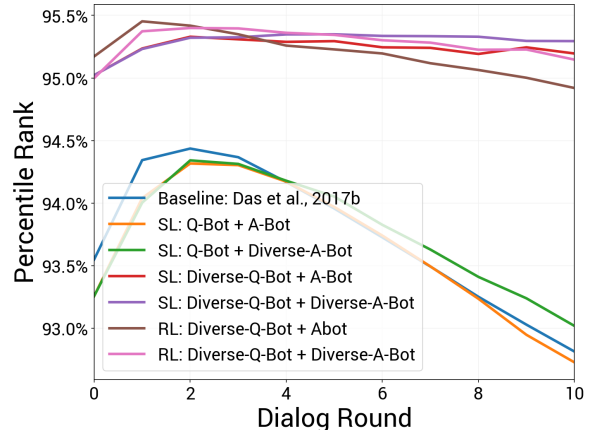


Figure 1: Performance on the image-guessing task. Percentile rank (higher is better) of the true image (shown to A-BOT) as retrieved using fc7 image feature predictions from Q-BOT.

every round from 1 through t , is concatenated with history representation from the previous round. This gives us a set of t question-history vectors for t rounds. These vectors are fed as input to a dialog-level LSTM, whose output state at t is used to decode the response to Q_t . 2 shows the model architecture of the HRE.

All LSTMs are 2-layered with 512-dim hidden states. We learn 300-dim embeddings for words and images. These word embeddings are shared across question, history, and decoder LSTMs.

References

- Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José M.F. Moura, Devi Parikh, and Dhruv Batra. 2017a. Visual Dialog. In *CVPR*. 1, 2, 3, 4
- Abhishek Das, Satwik Kottur, José M.F. Moura, Stefan Lee, and Dhruv Batra. 2017b. Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning. In *ICCV*. 2, 4

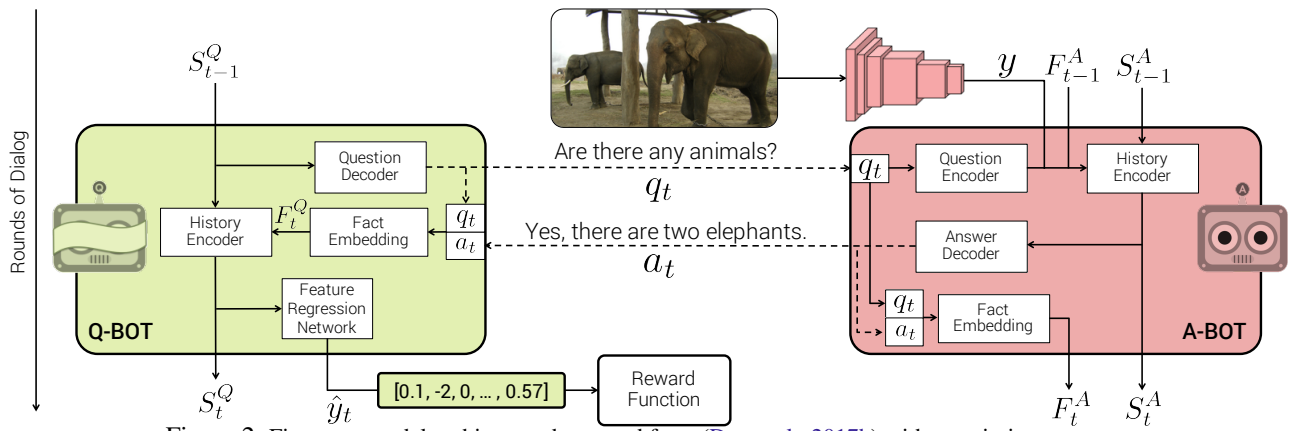


Figure 2: Figure on model architecture borrowed from (Das et al., 2017b) with permission.

	v1.0 val						v1.0 test-std					
	NDCG ↑	MRR ↑	R@1 ↑	R@5 ↑	R@10 ↑	Mean Rank ↓	NDCG ↑	MRR ↑	R@1 ↑	R@5 ↑	R@10 ↑	Mean Rank ↓
Baseline: Das et al. (2017b)	53.76	46.35	36.22	56.15	62.41	19.34	51.60	45.67	35.05	56.30	63.25	19.15
Diverse SL A-BOT	53.82	46.55	36.46	56.16	62.68	19.34	51.77	45.98	35.58	56.33	63.08	18.97
Diverse RL A-BOT (Finetuned with Diverse-A-BOT)	53.94	46.67	36.54	56.34	62.99	19.24	51.80	45.70	34.85	56.33	63.90	18.88

Table 1: A-BOT performance on the VisDial v1.0 dataset (Das et al., 2017a). Arrow indicates direction of better performance.

	Diversity						Relevance	
	# Novel questions ↑	# Unique questions ↑	Mutual overlap ↓	Ent-1 ↑	Ent-2 ↑	Dist-1 ↑	Dist-2 ↑	Negative log likelihood ↓
Baseline: Das et al. (2017b)	71	6.70 ± 0.07	0.58 ± 0.01	2.72 ± 0.01	3.03 ± 0.02	0.35 ± 0.0	0.43 ± 0.0	9.94
SL: Q-BOT + Diverse-A-BOT	53	6.58 ± 0.07	0.60 ± 0.01	2.71 ± 0.01	3.00 ± 0.01	0.34 ± 0.0	0.42 ± 0.0	10.05
SL: Diverse-Q-BOT + Diverse-A-BOT	136	7.42 ± 0.07	0.50 ± 0.01	2.81 ± 0.01	3.18 ± 0.01	0.38 ± 0.0	0.48 ± 0.0	10.10
RL: Diverse-Q-BOT + Diverse-A-BOT	288	7.87 ± 0.06	0.46 ± 0.01	2.86 ± 0.01	3.24 ± 0.01	0.39 ± 0.0	0.51 ± 0.0	10.32

Table 2: Q-BOT diversity and relevance on v1.0 val. ↑ indicates higher is better. ↓ indicates lower is better.

Image + Caption	Human-Human dialog (Das et al., 2017a)	Q-BOT-A-BOT dialog	Diverse-Q-BOT-A-BOT dialog
	<p>what color is the cutting board? blue</p> <p>how big is it about the size of a regular laptop screen</p> <p>how many carrots are there? 4</p> <p>are they the orange kind? 2 are orange, 2 are white</p> <p>what kind of cell phone is it? a samsung</p> <p>can you see any text on the screen? a recipe for UNK UNK soup</p> <p>do you see the room? just the counter</p> <p>are there people? not visible</p> <p>what color is the counter? white</p>	<p>is the photo in color? no</p> <p>how many people do you see? no people this a close up?</p> <p>no</p> <p>is this a close up? zebra is</p> <p>is this a close up? no</p> <p>is this a close up? no</p> <p>is this a close up? no</p> <p>is this a close up? no</p> <p>is this a close up? no</p> <p>is this a close up? no</p> <p>is this a close up? no</p>	<p>is the photo in color? 1 picture is in color</p> <p>is the photo close up? black and white</p> <p>is the photo in color? no it 's black and white</p> <p>is it a close up? no it 's a close up of the cake</p> <p>can you tell what brand it is? no i ca n't see the brand</p> <p>can you tell if it 's a male or female? no i ca n't tell</p> <p>can you tell if it 's a male or female? no i ca n't tell</p> <p>is this in a kitchen? nope it 's in a home</p> <p>are there any utensils? no</p> <p>are there any utensils? nope</p>
	<p>is the field large? yes</p> <p>is the grass green or brown? green</p> <p>can you see a barn? no</p> <p>can you see a fence? no</p> <p>can you see a water supply? yes</p> <p>can you see any people? no</p> <p>can you see trees? yes</p> <p>can you see mountains in the background? no</p> <p>is the grass tall or short? short</p>	<p>what color are the cows? they are brown and white</p> <p>how many cows are there? 4</p> <p>what color are the cows? they are brown and white</p> <p>can you see the sky? it is not visible</p> <p>do you see any trees? i do not see any trees</p> <p>can you see the sky? green sky</p> <p>are there any clouds? the sky is not visible</p> <p>how many cows are there? 2</p> <p>what color are the cows? the cows are brown and white</p> <p>what color are the cows? they are brown</p>	<p>do you see any people? 1</p> <p>do you see a fence? a small 1</p> <p>how many cows are there? 5</p> <p>what color are the cows? black and white</p> <p>can you see the sky? some</p> <p>can you see the sky? nope</p> <p>are there any other animals? not that i can see</p> <p>can you see the sky? not at all</p> <p>are there any trees? in the background</p> <p>are there any trees? not at all</p>
	<p>what type of building? may be part of a zoo</p> <p>how many elephants? about 4</p> <p>is it sunny? clear day</p> <p>any trees? yes</p> <p>what color is the building? dark grey</p> <p>are there other animals? no</p> <p>are there any people? yes</p> <p>what are they wearing? shorts and shirts</p> <p>how many? 3</p>	<p>how many elephants are there? ca n't tell</p> <p>what color are the elephants? ca n't tell</p> <p>how many elephants are there? ca n't tell</p> <p>what color is the fence? i ca n't tell</p> <p>can you see the sky? i ca n't see the sky</p> <p>are there any trees? just a few</p> <p>does this look like a zoo? 1 of them is</p> <p>are there any people? nothing in the picture</p> <p>how many elephants are there? just 1</p> <p>can you see the sky? just a little</p>	<p>how many elephants are there? 1</p> <p>is there any people? ca n't see any</p> <p>male or female? not sure</p> <p>just standing there</p> <p>male or female? there are 2 males and 1 female</p> <p>1 of it</p> <p>just a black and white picture</p> <p>there are trees in the background</p> <p>red and white</p> <p>i see some in the background</p>
	<p>is the photo in color? yes</p> <p>is there humans in the photo? no</p> <p>is there any animals in the photo? no</p> <p>what color are the walls? tan</p> <p>what color is the sink? white</p> <p>is there any towels? no</p> <p>is there any pictures on the wall? no</p> <p>what color is the tub? white</p> <p>can you see the floor? no</p>	<p>is the photo in color? 5 of them are</p> <p>are there any people? it 's a close up so i ca n't see any people</p> <p>what color is the sink? about the same color as the toilet</p> <p>do you see a mirror? no</p> <p>what color is the sink? yes it is white</p> <p>what color is the sink? yes it is white</p> <p>what shape is the mirror? yes, it 's a rectangle</p> <p>are there any towels? he 's on a towel rack</p> <p>are there any towels? i do n't see any towels</p> <p>are there any towels? i do n't see any towels</p>	<p>is the photo in color? no</p> <p>is there any people? 5</p> <p>what color walls? i ca n't see the walls</p> <p>is there a mirror? yes</p> <p>is there a mirror? yes</p> <p>is there a mirror? yes</p> <p>are the walls seen? no</p> <p>what color are the walls? no walls</p> <p>what is the floor made of? what color is the floor? i ca n't see the floor</p>
	<p>about how many cows? over a dozen</p> <p>are there any calves? no</p> <p>are they eating? yes</p> <p>is it sunny? no</p> <p>is there a fence around them? no</p> <p>do you see any buildings? no</p> <p>mountains in the background? no</p> <p>is there snow on the ground? no</p> <p>any trees? yes</p>	<p>is the photo in color? yes</p> <p>is there any people? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p>	<p>how many cows are there? about 10</p> <p>what color are the cows? black and white</p> <p>how many cows are there? about 10</p> <p>what color are the cows? black and white</p> <p>what is the weather like? sunny</p> <p>are there any trees? no</p> <p>are there any people? no</p> <p>are there any trees? no</p> <p>are there any buildings? no</p> <p>are there any trees? yes</p>
	<p>what color are the kitchen walls? white</p> <p>is there a window present? yes</p> <p>are there pictures paintings hanging up? no</p> <p>is the stove visible? yes</p> <p>what color is the stove? black and silver</p> <p>are there any people visible? no</p> <p>what color is the nearby staircase? dark brown</p> <p>is the staircase carpeted? no</p> <p>are the handrails of the staircase visible? yes</p>	<p>is the photo in color? yes</p> <p>is there any people? no</p> <p>is there a stove? no</p> <p>are there any windows? no</p> <p>can you see the floor? no</p> <p>can you see the floor? no</p> <p>can you see the floor? yes</p> <p>can you see the floor? yes</p> <p>can you see the floor? yes</p> <p>are the walls visible? yes</p>	<p>is the kitchen clean? brown</p> <p>is the kitchen clean? it 's cluttered</p> <p>are there any windows? yes</p> <p>is it daytime? yes</p> <p>are there any people? yes</p> <p>are the appliances stainless steel? i ca n't tell</p> <p>are the people in the kitchen? yes</p> <p>is the kitchen clean? yes</p> <p>are the people male or female? yes</p> <p>is the kitchen clean? yes</p>
	<p>how many men? 3</p> <p>how big is the kitchen? small industrial</p> <p>what is the age range? teens to 40s</p> <p>is the kitchen belong in a house or a restaurant? restaurant</p> <p>what are they doing in the kitchen? cooking and mixing food</p> <p>what kind of floor in the kitchen? i ca n't tell</p> <p>is there any decorations in the kitchen? no</p> <p>is the kitchen look clean? no</p> <p>does the men seem to be UNK? they seem UNK</p>	<p>can you see the fridge? the side of it</p> <p>what color is it? the photo is black and white</p> <p>can you see the sink? 1 of it</p> <p>can you see the sink? just a tiny bit of it</p> <p>can you see the sink? does n't look like it</p> <p>can you see the sink? 0</p> <p>can you see the sink? 4 of them</p> <p>can you see the fridge? UNK</p> <p>can you see the fridge? not real</p>	<p>how many men are there? looks like 2 men and 1 woman</p> <p>how old do they look? it 's hard to tell, but they look like they are in their 20s</p> <p>how old do they look? looks to be in their 30 's</p> <p>are they wearing aprons? just 1 of them is wearing hat</p> <p>are they wearing hats? it is hard to tell, but i think so</p> <p>can you see the stove? not at all</p> <p>are there any pots on the stove? not that i can see</p> <p>does the kitchen look clean? black and white photo but i think it 's clean</p> <p>can you see any windows? no</p> <p>do you see any windows? ca n't see any windows</p>
	<p>who is eating? is the drink in the cup soda</p> <p>what is the food on? bread and vegetable</p> <p>is it in a plate? what kind of sauces are there in the back</p> <p>i am asking the questions here? yes you are</p> <p>what is the food on? bread, vegetables, chicken and drink</p> <p>are they on a table? yes they are</p> <p>any people? there are n't any people in sight</p> <p>what color is the table? it is dark gray</p> <p>any cutlery? yes there is a butter knife in the plate</p>	<p>is the photo in color? it is</p> <p>is the photo close up? no</p> <p>is the photo in color? it is</p> <p>do you see any other people? it 's a close up of the man 's hand, so i ca n't tell</p> <p>how old is the man? maybe in his 30s</p> <p>is he wearing glasses? i ca n't see his face</p> <p>what color is his shirt? not sure, the photo is in black and white</p> <p>does he have facial hair? no</p> <p>does he wear glasses? yes</p>	<p>are there any people in the photo? yes</p> <p>is the photo in color? it is</p> <p>is it a professional photo? i do n't think so</p> <p>how many people? 0</p> <p>is there a tablecloth on the table? yes</p> <p>what color is the tablecloth? white</p> <p>are there any utensils visible? no</p> <p>what color is the plate? it 's white</p> <p>are there any utensils visible? no</p> <p>are there any utensils visible? no</p>

Table 3: Selected examples of dialog on v1.0 VisDial test split for different SL variants. We observe that variant involving Diverse-Q-BOT generates more diverse, image relevant and fluent dialog.

Image + Caption	Human-Human dialog (Das et al., 2017a)	(Das et al., 2017b) dialog	RL: Diverse-Q-BOT-A-BOT dialog
	<p>what color is the cutting board? blue</p> <p>how big is it about the size of a regular laptop screen</p> <p>how many carrots are there? 4</p> <p>are they the orange kind? 2 are orange, 2 are white</p> <p>what kind of cell phone is it? a samsung</p> <p>can you see any text on the screen? a recipe for UNK UNK soup</p> <p>do you see the room? just the counter</p> <p>are there people? not visible</p> <p>what color is the counter? white</p> <p>what color is the cell phone.</p>	<p>is the photo in color? no</p> <p>is the photo close up? no</p> <p>how many people? cows</p> <p>is this a close up? 2 of them are</p> <p>what are they wearing? no shirt and jeans</p> <p>what are they wearing? no shirt and jeans</p> <p>do you see other people? sunny 1 in the background</p> <p>what color are the scissors? no handle</p> <p>is this in a kitchen? yes</p> <p>do you see the counter? no</p>	<p>is the photo in color? black and white</p> <p>is there any people? I hand</p> <p>is this indoors or outdoors? no idea</p> <p>is the person male or female? up close, i ca n't tell</p> <p>can you tell what kind of sandwich it is? no</p> <p>is there a cutting board? no</p> <p>is there a knife? yes</p> <p>are the carrots cut up? daytime</p> <p>is the scene well lit? no</p> <p>are the carrots raw or cooked? short</p>
	<p>is the field large? yes</p> <p>is the grass green or brown? green</p> <p>can you see a barn? no</p> <p>can you see a fence? no</p> <p>can you see a water supply? yes</p> <p>can you see any people? no</p> <p>can you see trees? yes</p> <p>can you see mountains in the background? no</p> <p>is the grass tall or short? short</p>	<p>can you see any people? nope</p> <p>can you see the sky? a little bit</p> <p>i ca n't see the sky? i ca n't see the sky</p> <p>are there any clouds? not that i can see</p> <p>are there any other animals? lots of trees</p> <p>do you see any trees? ca n't tell, they are far away</p> <p>how many cows are there? i do not</p> <p>do you see any buildings? i do not</p> <p>can you see the sky? ca n't see the sky</p> <p>do you see any clouds? ca n't see the sky</p> <p>can you see the sky? ca n't see the sky</p>	<p>are the cows all the same color? they are different shades of brown</p> <p>do any of the cows have horns? it does n't look like it</p> <p>do you see a fence? there is no fence</p> <p>how many cows are there? 4</p> <p>can you see the sky? 1 of it</p> <p>what color is the grass? the grass is green</p> <p>can you see the sky? it is not visible</p> <p>does it appear to be sunny? not really</p> <p>can you see the sky? not at all</p> <p>can you see the ground? a little bit</p>
	<p>what type of building? may be part of a zoo</p> <p>how many elephants? about 4</p> <p>is it sunny? clear day, yes</p> <p>any trees? yes</p> <p>what color is the building? dark grey</p> <p>are there other animals? no</p> <p>are there any people? yes</p> <p>what are they wearing? shorts and shirts</p> <p>how many? 3</p>	<p>do you see people? 2</p> <p>how many elephants? not sure</p> <p>does the elephant have tusks? not that i can see</p> <p>does the elephant have tusks? in the distance, yes</p> <p>does the elephant have tusks? not that i can see</p> <p>any other animals? not that i can see</p> <p>does the elephant have tusks? not that i can see</p> <p>does the elephant have tusks? not that i can see</p> <p>does the elephant have tusks? not that i can see</p> <p>can you see the sky? 1 of it</p>	<p>can you see any people? yes</p> <p>what are the elephants doing? just standing around</p> <p>how many people are there? a lot, too many to count</p> <p>does this look like a zoo? no</p> <p>how many elephants are there? in the distance there are many</p> <p>does this look like a parade? a little bit</p> <p>does this look like a circus? nope</p> <p>can you see any signs? there is a sign on the street</p> <p>does it look like it is in the usa? i do n't think so</p> <p>can you see the sky? i can</p>
	<p>is the photo in color? yes</p> <p>is there humans in the photo? no</p> <p>is there any animals in the photo? no</p> <p>what color are the walls? tan</p> <p>what color is the sink? white</p> <p>is there any towels? no</p> <p>is there any pictures on the wall? no</p> <p>what color is the tub? white</p> <p>can you see the floor? no</p>	<p>how big is the bathroom? ca n't see the whole room</p> <p>how big is the mirror? i can only see a small part of it</p> <p>what color is the sink? there is no sink</p> <p>what color is the sink? black</p> <p>what color is the sink? yes</p> <p>what color are the walls? he is white</p> <p>is there a mirror? yes</p> <p>are there any people? he is alone</p> <p>are there any animals? white cat</p> <p>what color is the sink? yes</p>	<p>how big is the bathroom? it 's pretty big</p> <p>is it clean? maybe</p> <p>how big is the mirror? i can only see a small part of it</p> <p>what color is the wall? maybe white</p> <p>is there a shower curtain? there is no shower curtain</p> <p>is there a mirror? yes</p> <p>what shape is the mirror? yes, it 's a rectangle</p> <p>what color is the wall? i ca n't see the wall</p> <p>are there any people? he 's the only 1</p> <p>is there a window? i do n't see 1</p>
	<p>about how many cows? over a dozen</p> <p>are there any calves? no</p> <p>are they eating? yes</p> <p>is it sunny? no</p> <p>is there a fence around them? no</p> <p>do you see any buildings? no</p> <p>are there any mountains in the background? no</p> <p>is there snow on the ground? no</p> <p>any trees? yes</p>	<p>is the photo in color? yes</p> <p>is there any people? no</p> <p>how many cows? black and white cows</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p> <p>is there a fence? no</p>	<p>how many cows are there? about 10</p> <p>what color are the cows? black and white</p> <p>are any of them babies? no</p> <p>how many cows are there? about 10</p> <p>are there any animals in the photo? no</p> <p>are there any other animals in the photo? no</p> <p>are there any buildings in the photo? no</p> <p>are there any trees? yes</p> <p>is there a fence? no</p> <p>is it sunny out? yes</p>
	<p>what color are the kitchen walls? white</p> <p>is there a window visible? yes</p> <p>are there pictures paintings hanging up? yes</p> <p>is the stove visible? yes</p> <p>what color is the stove? black and silver</p> <p>are there any people visible? no</p> <p>what color is the nearby staircase? dark brown</p> <p>is the staircase carpeted? no</p> <p>are the handrails of the staircase visible? yes</p>	<p>is this in color? no</p> <p>do you see any people? no</p> <p>do you see a stove? standing</p> <p>are there any windows? white windows</p> <p>can you see the floor? metal floor</p> <p>can you see the floor? brown tile</p> <p>can you see the floor? yes</p> <p>is it tiled? no</p> <p>do you see a window? short window</p> <p>are there curtains on the window? yes</p>	<p>what color are the appliances? black and white</p> <p>is there a window? yes</p> <p>is there a sink? yes</p> <p>can you see a refrigerator? yes</p> <p>are there any people? yes</p> <p>is it daytime? nope</p> <p>is it sunny out? just a little</p> <p>is there a window? brown</p> <p>can you see the floor? no</p> <p>do you see a stove? no</p>
	<p>how many men? 3</p> <p>how big is the kitchen? small industrial</p> <p>what is the age range? teens to 40s</p> <p>is the kitchen belong in a house or a restaurant? restaurant</p> <p>what are they doing in the kitchen? cooking and mixing food</p> <p>what kind of floor in the kitchen? i ca n't tell</p> <p>is there any decorations in the kitchen? no</p> <p>is the kitchen look clean? no</p> <p>does the men seem to be UNK? they seem UNK</p>	<p>can you see the fridge? yep</p> <p>can you see the fridge? kind of</p> <p>does it have a freezer on it? nope</p> <p>how many men are there? ca n't tell</p> <p>how old do they look? it 's hard to tell, but i 'd say in their 30 's</p> <p>what color are the walls? it 's a black and white photo so i ca n't tell</p> <p>how old are the men? i ca n't see their faces, i do n't know</p> <p>does the kitchen look clean? 0</p> <p>how old do the men look? i ca n't see their faces, i do n't know</p> <p>are they wearing aprons? a few are</p>	<p>can you tell what kind of food they are? i can not tell what kind of food it is</p> <p>does the kitchen look clean? fairly clean</p> <p>do you see any windows? partially in the background</p> <p>do you see any curtains? maybe 1 in the background</p> <p>is it day or night? ca n't tell</p> <p>are the men old or young? ca n't see their faces</p> <p>are they wearing hats? just 1</p> <p>are they wearing glasses? it looks like they are</p> <p>are they smiling? i ca n't see their faces</p> <p>are they smiling? i ca n't see their faces</p>
	<p>who is eating? is the drink in the cup soda</p> <p>what is the food on? bread and vegetable</p> <p>is it in a plate? what kind of sauces are there in the back</p> <p>i am asking the questions here? yes you are</p> <p>what is the food on? bread, vegetables, chicken and drink</p> <p>are they on a table? yes they are</p> <p>any people? there are n't any people in sight</p> <p>what color is the table? it is dark gray</p> <p>any cutlery? yes there is a butter knife in the plate</p>	<p>are there any people? i see part of a person 's hand</p> <p>what color is the plate? black and white</p> <p>are there any utensils? no</p> <p>are there any utensils? no</p> <p>are there any utensils? silver spoon</p> <p>are there any utensils? i see a fork</p> <p>are there any utensils? yes a fork and a knife</p> <p>are there any napkins? 0 that i can see</p> <p>are there any napkins? it 's a close up so i ca n't tell</p>	<p>are there any people in the picture? yes</p> <p>are they inside or outside? outside</p> <p>what gender are the people? i ca n't tell</p> <p>what are the people doing? I is looking at the camera, the other is looking at the camera</p> <p>can you tell if they are male or female? i ca n't tell</p> <p>can you tell if the person is male or female? i ca n't see their face, i do n't know</p> <p>can you tell if they are male or female? i ca n't tell</p> <p>can you tell what kind of sandwich it is? it has some meat on it, but i ca n't tell what it is</p> <p>is it on a plate? no</p> <p>what color is the handle of the knife? silver</p>

Table 4: Selected examples of dialog on v1.0 VisDial test split for different RL variants. We observe that variant involving Diverse-Q-BOT generates more diverse, image relevant and fluent dialog.