A Appendices

A.1 Data Statistics

As shown in Table 5, there are 30 *(domain, slot)* pairs on MultiWOZ (Budzianowski et al., 2018) dataset, including 10 slots in the *hotel* domain, 6 slots in the *train* domain, 3 slots in the *attraction* domain, 7 slots in the *restaurant* domain, and 4 slots in the *taxi* domain. *hotel* and *restaurant* domains, in particular, have the most dialogues in the dataset. *Taxi* domain, on the other hand, is the one with less data.

	Hotel	Train	Attraction	Restaurant	Taxi
Slots	price, type, parking, stay, day, people, area, stars, internet, name	destination, departure, day, arrive by, leave at, people	area, name, type	food, price, area, name, time, day, people	destination, departure, arrive by, leave at
Train	3381	3103	2717	3813	1654
Dev	416	484	401	438	207
Test	394	494	395	437	195

Table 5: In total, there are 30 *(domain, slot)* pairs from the selected five domains on MultiWOZ. The numbers in the last three rows indicate the number of dialogues.

Figure 6: Attention Visualization

Figure 7: Attention Visualization

	1%	**5%**	**10%**	**25%**
TRADE (w/o Ont.)	10.35 (12.58)	27.70 (31.17)	32.61 (36.18)	38.46 (42.71)
+ Consistency	13.77 (15.58)	27.95 (31.89)	33.64 (37.40)	40.06 (43.45)
+ Behavior	18.69 (22.10)	29.95 (34.30)	**34.57 (38.43)**	39.42 (42.45)
Consistency + Behavior	**19.50 (21.90)**	**30.59 (35.13)**	34.50 (38.12)	**40.15 (43.40)**

Table 6: Semi-supervised learning joint goal accuracy (and its fuzzy matching version in parentheses) on MultiWOZ-2.1 test set from 1% to 25% training data.

Figure 8: Attention Visualization

Figure 9: Attention Visualization